

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 9, 2003, 19:24:17 ; Search time 2554 Seconds
(without alignments)
10608.868 Million cell updates/sec

Title: US-09-997-900-1
Perfect score: 1673
Sequence: 1 gctcttcagtagcaaaaa.....tgatttgggtcatttcg 1673

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	565.4	33.8	1380	11	AY105043
2	502.4	30.0	539	9	AI992657
3	494.2	29.5	605	10	AV439911
4	404.6	24.2	508	14	N96256
5	399	23.8	423	10	AV801585
6	392	23.4	802	14	B0515774

Result No.	Score	Query Match	Length	DB ID	Description
7	376.8	22.5	454	14	T88501
8	372.4	22.3	418	10	AV796666
9	366.4	21.9	679	14	B0407261
10	366	21.9	714	14	B0863607
11	359.2	21.5	431	10	AV797506
12	346.2	20.7	422	10	AV816962
13	338.2	20.2	410	10	AV805399
14	334.6	20.0	607	10	AW10049
15	333.2	19.9	675	10	AM099716
16	332	19.8	631	10	AM585871
17	328.2	19.6	508	9	AA394708
18	325.6	19.5	782	12	BG581436
19	323.8	19.4	541	12	BE995444
20	322	19.2	618	14	B0913769
21	322	19.2	797	12	BG36929
22	321.6	19.2	598	10	AM508667
23	311	18.6	504	12	BG507593
24	310.6	18.6	632	14	BQ415906
25	309.8	18.5	638	12	BE823545
26	299.6	17.9	409	10	AV808789
27	296.4	17.7	601	10	AM696137
28	296	17.7	624	10	AM181183
29	294.4	17.6	688	13	BJ257372
30	288.8	17.3	552	13	BJ310586
31	283	16.9	382	10	AV790195
32	282	16.9	674	13	BI931584
33	279	16.7	555	10	BE433354
34	277.2	16.6	618	10	AV916746
35	275.2	16.4	653	12	BF481461
36	272.6	16.3	388	10	AV804485
37	272.2	16.3	631	13	BE346287
38	272	16.3	466	10	BE346287
39	266.2	15.9	395	10	AV817239
40	266	15.9	417	17	BH847386
41	262.6	15.7	517	12	BF068113
42	258.8	15.5	660	10	AV922042
43	255.8	15.3	505	14	BQ629151
44	255.6	15.3	722	12	BG126660
45	252.6	15.1	795	14	BQ515775

ALIGNMENTS

RESULT 1
LOCUS AY105043 1380 bp mRNA linear HTC 25-MAY-2002
DEFINITION Zea mays
ACCESSION AY105043
VERSION AY105043.1 GI:21208121
KEYWORDS
SOURCE
ORGANISM Zea mays.
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PAC clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 1380)
Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
Maize Mapping Project/Dupont Consensus Sequences for Design of Overgo Probes
JOURNAL Unpublished (2002)
REFERENCE 2 (bases 1 to 1380)
AUTHORS Coe,E.C.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
FEATURES
source 1..1380
/organism="Zea mays"
/db_xref="MaizeDB:633742"
/db_xref="taxon:4577"
/clone="PC0069007"


```
Db 420 GAAGTCATCTGATTAGATTGCCGTGAACGCTGCTGTAGAGAAGATGTCCTGGACATT 361
QY 1305 GCTAGTATTTTCAGGCGCTAAAGCTGTTGACGTATTCATCACACAATTTACTTGGACGTT 1364
Db 360 GCATGATTTTTCAGGCGCTAAAGCTGTTGACGTATTCATCACACAATTTACTTGGACGTT 301
QY 1365 ACTGGGATCTAGACAAAGATGTTGGACACTGCAAAAGTTATTGGAGCCCTATGTTATATGT 1424
Db 300 ACTGGGATCTAGACAAAGATGTTGGACACTGCAAAAGTTATTGGAGCCCTATGTTATATGT 241
QY 1425 GAGGTTCGAAGAACCCTGCTGTGCGATTGGCTGCTGATCGGAGTGGACTCCAACTAC 1484
Db 240 GAGGTTCGAAGAACCCTGCTGTGCGATTGGCTGCTGATCGGAGTGGACTCCAACTAC 181
QY 1485 CTTCGTGATCTCTCTTCTTTTACAGGCTAACCGTTGCAGAGTCATCCATCGACACA 1544
Db 180 CTTCGTGATCTCTCTTCTTTTACAGGCTAACCGTTGCAGAGTCATCCATCGACACA 121
QY 1545 TCAGAACTTTGGAAGCTTAAAGCTTTCATTACACAGTCTATGAAAC--CTCAAGACAGA 1601
Db 120 TCAGAACTTTGGAAGCTTAAAGCTTTCATTACACAGTCTATGAAACCTCAAGACAGA 61
QY 1602 CAGAGAGACTGCTCATATATGTTGTGACTTTGTTTATGAACAATTTAGCTGATTTTG 1661
Db 60 CAGAGAGACTGCTCATATATGTTGTGACTTTGTTTATGAACAATTTAGCTGATTTTG 1

RESULT 3
AV439911/c 605 bp mRNA linear EST 14-NOV-2000
LOCUS AV439911 Arabidopsis thaliana above-ground organ two to six-week
DEFINITION Old Arabidopsis thaliana cDNA clone APD28a06_f_3', mRNA sequence.
ACCESSION AV439911
VERSION AV439911.1 GI:7610260
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 605)
Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)

JOURNAL
MEDLINE 20363093
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
source
1..605
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="APD28a06_f"
/clone_lib="Arabidopsis thaliana above-ground organ two to
six-week old"
/tissue_type="aboveground organs"
/dev_stage="two to six-week old"
/note="Vector: pBluescriptII SK-, Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 193 a 135 c 110 g 167 t
ORIGIN
Query Match 29.5%; Score 494.2; DB 10: Length 605;
Best Local Similarity 98.8%; Pred. No. 1,2e-142;
Matches 509; Conservative 0; Mismatches 3; Indels 3; Gaps 1;
QY 1151 AGATGAATCGGTGACGAATTTGGTGCAGCACTTACAAACGTAGATGTGATGAGGT 1210
|||||
```

```
Db 605 AGATGAATCGGTGACGAATTTGGTGCAGCACTTACAAACGTAGATGTGATGAGGT 546
QY 1211 CCATGATCTTACCATTTGGCATTTTCTGAAAGAGACGTGCTGATTAGATGGCGT 1270
Db 545 CCATGATCTTACCATTTGGCATTTTCTGAAAGAGACGTGCTGATTAGATGGCGT 486
QY 1271 GAACGCTGCTGATAGAGAAGATGCTCGACATTGCTAGTATTTTTCAGGCTAAAGCTGT 1330
Db 485 GAACGCTGCTGATAGAGAAGATGCTCGACATTGCTAGTATTTTTCAGGCTAAAGCTGT 426
QY 1331 TGACGTATCTGATCACACAATTTACTTTGACGTTACTGGGACTGTAGACAAATGTTGC 1390
Db 425 TGACGTATCTGATCACACAATTTACTTTGACGTTACTGGGACTGTAGACAAATGTTGC 366
QY 1391 ACGCAAGGTTATTGGAGCCCTATGCTATGAGGTGCAAGAACCGCTGTGGCG 1450
Db 365 ACGCAAGGTTATTGGAGCCCTATGCTATGAGGTGCAAGAACCGCTGTGGCG 306
QY 1451 ATTGGCTGCTGAATCGGAGTGCATCCAAAGTACCTTCGTGATCTCTTTCTTTTAC 1510
Db 305 ATTGGCTGCTGAATCGGAGTGCATCCAAAGTACCTTCGTGATCTCTTTCTTTTAC 246
QY 1511 AGCTTAACCGTTGCAGAGTGCATCCATGCATCAGAAACTTTGGAGAGTTAAAGTTT 1570
Db 245 AGCTTAACCGTTGCAGAGTGCATCCATGCATCAGAAACTTTGGAGAGTTAAAGTTT 186
QY 1571 CATTACACAGTCTATGAAAC--CTCAAGACAGACAGAGACTGCTCATATGTTT 1627
Db 185 CATTACACAGTCTATGAAACCTCAAGACAGACAGAGACTGCTCATATATGTTT 126
QY 1628 GTGACTTTGTTTATGAACAATTTAGCTGATTTTG 1662
Db 125 GTGACTTTGTTTATGAACAATTTAGCTGATTTTG 91

RESULT 4
N96256 508 bp mRNA linear EST 05-JAN-1998
LOCUS N96256
DEFINITION 21856 Lambda-PRL1 Arabidopsis thaliana cDNA clone G7C11T7, mRNA
sequence.
ACCESSION N96256
VERSION N96256.1 GI:2748591
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 508)
Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh
L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M., Retzel
E. and Somerville,C.
Genes galore: a summary of methods for accessing results from
large-scale partial sequencing of anonymous Arabidopsis cDNA clones
Plant Physiol. 106, 1241-1255 (1994)
95148729
On Jan 5, 1998 this sequence version replaced g1:1269029.
Contact: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
Lansing, MI
Tel: 517-353-0854
Fax: 517-353-9168
Email: 22313tcn@lpm.cl.msu.edu
Seq primer: T7.
Location/Qualifiers
1..508
/organism="Arabidopsis thaliana"
/strain="var Columbia"
/db_xref="taxon:3702"
/clone="G7C11T7"
/clone_lib="Lambda-PRL1"
```


same plants as 3 but aerial tissue (stems, flowers and siliques). The vector is BRL's lambda Zip-Lox. The cDNA inserts were directionally cloned with Sal-Not arms using oligo dT primed cDNA.

BASE COUNT 124 a 92 c 104 g 118 t 16 others

ORIGIN

Query Match 22.5%; Score 376.8; DB 14; Length 454;
Best Local Similarity 91.2%; Pred. No. 5.1e-106;
Matches 412; Conservative 0; Mismatches 38; Indels 2; Gaps 2;

QY 728 AAGACAGCAAAATATAGCACTGGAAGGGAAGAAAGATGGCTGCACTGCTCCATTTGGCG 787
DB 1 AAGCAGGNAAGATAGCACTGGAAGGGAAGAAAGATGGCTGCACTGCTCCATTTGGCG 60
QY 788 ATTTCAGCAGCATTCATCCATGATCTCAAGAGCAAGCGCTGTAGTGTCTTGAGAG 847
DB 61 ATTTTNGCAGCATTCATCCATGATCTCAAGAGCAAGCGCTGTAGTGTCTTGAGAG 120
QY 848 TAGCAAAAAGAGCACTGCTCCCTCAAAAAGAAACATGACGAGGAGATGTTATCC 907
DB 121 TAGCAAAAAGAGCACTGCTCCCTCAAAAAGAAACATGACGAGGAGATGTTATCC 180
QY 908 CGTTGAGCCATTTTGGACCCCAAGTACATGCTATTTCTGACGCTGAGGACTTCT 967
DB 181 CGTTGAGCCATTTTGGACCCCAAGTACATGCTATTTCTGACGCTGAGGACTTCT 240
QY 968 CACTGAGAAATAGCAGTGGACTAGCGCTGCTACTATCTATCTGCTGTAATGATAT 1027
DB 241 CACTGAGAAATAGCAGTGGACTAGCGCTGCTACTATCTATCTGCTGTAATGATAT 300
QY 1028 TCCAGAGATCTTAATATGTGACTGCTGCTTTCCTCGAAGGAGATACAAATCCAGAG 1087
DB 301 TCCAGAGATCTTAATATGTGACTGCTGCTTTCCTCGAAGGAGATACAAATCCAGAG 360
QY 1088 CTGGCCGATGAGCATGCTGAAACCAAGGCAATTCACGATTAACAAGTTATACCTGC 1147
DB 361 CTGGCCGATGAGCATGCTGAAACCAAGGCAATTCACGATTAACAAGTTATACCTGC 419
QY 1148 AACGATGAATCGCTGAGCAAAATTTGGTCAGC 1179
DB 420 AACGATGAATCGCTGAGCAAAATTTGGTCAGC 450

RESULT 8
AV796666/c 418 bp mRNA linear EST 29-MAR-2002
LOCUS AV796666 RAFL9 Arabidopsis thaliana cDNA clone RAFL09-09-C24 3'

DEFINITION mRNA sequence.
ACCESSION AV796666
VERSION AV796666.1 GI:19830649
KEYWORDS EST.

SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 418)
Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Saccu, M., Nakajima, M.,
Oono, Y., Sekurai, T., Carninci, P., Kawai, J., Itoh, M., Ishi, Y.,
Arai, K., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.,
and Shinozaki, K.
Large scale analysis of Arabidopsis full-length cDNA (2002b)
Unpublished (2002)

TITLE JOURNAL COMMENT
Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@rct.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998). cDNA cleaved with BamHI

and XhoI was ligated to modified lambda FLX-1 vector (Carninci et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified Bluescript vector. Please visit our web site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further details.

FEATURES
source location/Qualifiers
1. 418
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone="RAFL09-09-C24"
/clone_11b="RAFL9"
/dev_stage="plants at various developmental stages from germination to mature seeds"
/lab_host="DH10B"
/note="Site 1: BamHI; Site 2: SalI; subjected to dehydration (1, 2, 5, 10, 24 hr) treatments"

BASE COUNT 126 a 96 c 81 g 115 t

ORIGIN

Query Match 22.3%; Score 372.4; DB 10; Length 418;
Best Local Similarity 96.4%; Pred. No. 1.1e-104;
Matches 403; Conservative 0; Mismatches 11; Indels 4; Gaps 2;

QY 1251 ATGCTGATTAGATTGCGCTGAAACGCTGCTGTAAGAAGATGCTGGACATTTGCTAGT 1910
DB 418 ATGCTGATTAGATTGCGCTGAAACGCTGCTGTAAGAAGATGCTGGACATTTGCTAGT 359
QY 1311 ATTTTGAAGGCTTAA-GCTGTTGACCTATCTGATACACAAATTTACTTTCAGCTTACTGG 1369
DB 358 ATTTTGAAGGCTTAAAGGCTGTTGACCTATCTGATACACAAATTTACTTTCAGCTTACTGG 299
QY 1370 GGATCTAGCAAGATGCTTGAAGGCTGCAAGGTTATGAGCCCTATGATATGAGGT 1429
DB 298 GGATCTAGCAAGATGCTTGAAGGCTGCAAGGTTATGAGCCCTATGATATGAGGT 239
QY 1430 TGCAGAACCGGTCGTGTGGCATTTGGCTGTAATCGGAGTGGACTCCAGTACTCTTG 1489
DB 238 TGCAGAACCGGTCGTGTGGCATTTGGCTGTAATCGGAGTGGACTCCAGTACTCTTG 179
QY 1490 TGGATATCTCTTTCTTTTAAACAGGCTTAAACCGTTGACAGTGCATCATCAATCAGA 1549
DB 178 TGGATATCTCTTTCTTTTAAACAGGCTTAAACCGTTGACAGTGCATCATCAATCAGA 119
QY 1550 AACTTGAAGGTAAGATTTTCAATTCACAGTCTATGAAC---CTCAAGACAGACAGAG 1606
DB 118 AACTTGAAGGTAAGATTTTCAATTCACAGTCTATGAACCACTCAAGACAGACAGAG 59
QY 1607 AGACTGCGTGAATATATGTTTGTGACTTTGTTATGAACAATTTAGCTGATTTTGGCC 1664
DB 58 AGACTGCGTGAATATATGTTTGTGACTTTGTTATGAACAATTTAGCTGATTTTGGCC 1

RESULT 9
BQ407261 679 bp mRNA linear EST 22-MAY-2002
LOCUS BQ407261 GA_E0104F04f Gossypium arboreum 7-10 dpa fiber library Gossypium

DEFINITION GA_E0104F04f Gossypium arboreum 7-10 dpa fiber library Gossypium
arborescens cDNA clone GA_E0104F04f, mRNA sequence.

ACCESSION BQ407261
VERSION BQ407261.1 GI:21094948
KEYWORDS EST.

SOURCE Gossypium arboreum.
ORGANISM Gossypium arboreum.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.

REFERENCE 1 (bases 1 to 679)
Wing, R.A., Fritsch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry

D., Wood, T.C., Leslie, A. and Wilkins, T.A.
An integrated analysis of the genetics, development, and evolution
of the cotton fiber
Unpublished (2000)
Contact: Wing RA

Db 121 ATGTTTAAAGTTGAGATATCTCAAGGACACAGTAGAAGTGAATGATGCTCA 180
553 TAAAGTAGATGACATCCAGATCCAGGAGAGATCATGTGCTAGTAGACATTC 612
181 TTAAGATCATGAGATCCAGAGTACCGATGAGGTTAGTGTGGTGGACATCTTA 240
613 GACCAAGATGTAGATATGAGGAGACATGCTGACTATCGAGTAACATGAGATCCTG 672
241 GAGCTAGATGTTGATATTTTCAAGAGACTATTAACCATGTAAGTAACTGGTACCTG 300
673 GAAAGATGATGCTGTGAGAAAGATTTGAAAAAGTTTCAGATCAGAGATTTGAAGA 732
301 GAAAGATGTTGAGTGCAGAAAGATTTGAAGCAAGTTGGAAATCCGAGATTGCAAGGA 360
733 CAGGAAAGATGACATGAGAGAGGAGAAAGATGGTGCACCTGCTCATTTTGGCGATTTT 792
361 CTGGAAAGATGCTGCTTGAGAGAGAAAAATGAGTGAATTCCTCTTTTGGAGATTTT 420
793 CAGCAGCATCTTATCCAGATCTCAGAGAGCAGCGCTGTTAGTGTCTTTCGAAAGTAGCA 852
421 CAGCAGCTTATATCCAGAGC-----TTGAAGCTCCTCTTCAGACCA 462
853 AAAAGAGAGCATGTCCTCCCAAGAAACATCAGCAGGAGAGATGTTTATCCGCTTG 912
463 AAACAAAAACGAAATATCTGAGATCTGATACATGATGGGAGAGATGTTATCCAGTTG 522
913 AGCCATTTTGAACCCCAAGTACATGATCTGATCTGCAGCTCAGCGGAGCTCTCAGTG 972
523 AGCTATATGATGACTATGCTTCAATTCAGTCTTGTGATGACATTTGGGCTGTTCTTAACG 582
973 AGCAAGATACGATGAGATACAGGCTGCATATCTATCATTTGCTTAAATGATATTTCCAG 1032
583 AAGAGATACACCGGCTCTGCTCACACACTTATCAATGATTTGAAGAGACTCTCCG 642
1033 GAATTTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1092
643 GAATTCATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 702
QY 1093 CCCTAGACATG 1104
Db 703 CAGTTGTCTATG 714

RESULT 11
AV797506/c 431 bp mRNA linear EST 29-MAR-2002
LOCUS AV797506 RAFL9 Arabidopsis thaliana cDNA clone RAFL09-12-G04 3',
DEFINITION mRNA sequence.
ACCESSION AV797506
VERSION AV797506.1 GI:19831489
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 431)
Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M.,
Oono, Y., Sekurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y.,
Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.
and Shinozaki, K.
Large scale analysis of Arabidopsis full-length cDNA (2002b)
Unpublished (2002)
Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@fc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
and XhoI was ligated to modified Lambda FL-C-1 vector (Carninci et

al., submitted for publication) digested with BamHI and SalI. This
clone is in a modified Bluescript vector. Please visit our web
site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further
details.

FEATURES
source location/Qualifiers
1..431
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone="RAFL09-12-G04"
/clone_lib="RAFL9"
/dev_stage="plants at various developmental stages from
germination to mature seeds"
/lab_host="DH10B"
/note="Site 1: BamHI; Site 2: SalI; subjected to
dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24
hr) treatments"

BASE COUNT 136 a 100 c 75 g 120 t
ORIGIN

Query Match 21.5%; Score 359.2; DB 10; Length 431;
Best local similarity 98.4%; Pred. No. 1.6e-100;
Matches 374; Conservative 0; Mismatches 3; Indels 3; Gaps 1;

QY 1286 AAGAGATGCTGAGCATTTGATGATTTTCAAGGCTAAAGCTTTGAGTATGATCA 1345
Db 431 AAGAGATGCTGAGCATTTGATGATTTTCAAGGCTAAAGCTTTGAGTATGATCA 372
QY 1346 CACATTTACTTTGACGCTTACTGAGGATCTAGACAAAGATGTTGACCTCAAGCTTATT 1405
Db 371 CACAAATTTACTTTGACGCTTACTGAGGATCTAGACAAAGATGTTGACCTCAAGGCTTATT 312
QY 1406 GGAGGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1465
Db 311 GGAGGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 252
QY 1466 GGGAGTGAATCCAGTACCTTCTGATGATGATGATGATGATGATGATGATGATGATG 1525
Db 251 GGGAGTGAATCCAGTACCTTCTGATGATGATGATGATGATGATGATGATGATGATGATG 192
QY 1526 AGAGTGCATTCATCGAATCAGAACTTTGGAAGGTAAGATTTTATACACAGCTCTAT 1585
Db 191 AGAGTGCATTCATCGAATCAGAACTTTGGAAGGTAAGATTTTATACACAGCTCTAT 132
QY 1586 GAAC--CTCAAGACAGACAGAGACTGCGTATATATGTTTGTGACTTTGTTATG 1642
Db 131 GAACCACTCAAGACAGACAGAGACTGCGTATATATGTTTGTGACTTTGTTATG 72
QY 1643 AAACATTTAGCTGATTTTGG 1662
Db 71 AAACATTTAGCTGATTTTGG 52

RESULT 12
AV816962/c 422 bp mRNA linear EST 01-APR-2002
LOCUS AV816962 RAFL9 Arabidopsis thaliana cDNA clone RAFL09-92-K10 3',
DEFINITION mRNA sequence.
ACCESSION AV816962
VERSION AV816962.1 GI:19858828
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 422)
Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M.,
Oono, Y., Sekurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y.,
Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.
and Shinozaki, K.
Large scale analysis of Arabidopsis full-length cDNA (2002b)
Unpublished (2002)
Contact: Motoaki Seki

Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060

Email: meskierc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998). cDNA cleaved with BamHI and XhoI was ligated to modified lambda FLc-1 vector (Carninci et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified Bluescript vector. Please visit our web site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further details.

FEATURES

source Location/Qualifiers

1..422
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone="RAFL09-92-K10"
/dev_stage="plants at various developmental stages from germination to mature seeds"
/lab_host="DH10B"
/note="Site:1: BamHI; Site:2: SalI; subjected to dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24 hr) treatments"
BASE COUNT 135 a 96 c 73 g 118 t
ORIGIN

Query Match 20.7%; Score 346.2; DB 10; Length 422;
Best Local Similarity 98.4%; Pred. No. 1.8e-96;
Matches 361; Conservative 0; Mismatches 3; Indels 3; Gaps 1;

QY 1299 GACATTGCTAGTATTTCAGGCGCTAAAGCTGTTGACGTATCGATCAGACAAATTACTTTG 1358
DB 422 GACATTGCTAGTATTTCAGGCGCTAAAGCTGTTGACGTATCGATCAGACAAATTACTTTG 363
QY 1339 CAGCTTACTGGGATCTAGACAAAGATGTTGCAGCTGCAAAAGTTATTGGAGCCCTATGCT 1418
DB 362 CAGCTTACTGGGATCTAGACAAAGATGTTGCAGCTGCAAAAGTTATTGGAGCCCTATGCT 303
QY 1419 AATATGTGAGTTGCAAGAACCGGTCGTGTGCGATTGGCTCGTGAAATCGGAGTGCATCC 1478
DB 302 AATATGTGAGTTGCAAGAACCGGTCGTGTGCGATTGGCTCGTGAAATCGGAGTGCATCC 243
QY 1479 AAGTACCTTCGTGAACTCCTTTCTTTTAACAGGCTAAACGGTTGCAGAGTCATCCAT 1538
DB 242 AAGTACCTTCGTGAACTCCTTTCTTTTAACAGGCTAAACGGTTGCAGAGTCATCCAT 183
QY 1539 CGAATCATGAAACTTTGGAAGGTAAAGTTTCATTACACAGTCTATGAAAC--CTCAAA 1595
DB 182 CGAATCATGAAACTTTGGAAGGTAAAGTTTCATTACACAGTCTATGAAACCAACTCAAA 123
QY 1596 GACACACAGAGAGCTGGCTCATATATGTTGTGACTTTGTTTATGAACAATTAGCTG 1655
DB 122 GACACACAGAGAGCTGGCTCATATATGTTGTGACTTTGTTTATGAACAATTAGCTG 63
QY 1656 ATTTTGG 1662
DB 62 ATTTTGG 56

RESULT 13

AV805399/c

LOCUS AV805399 410 bp mRNA linear EST 29-MAR-2002
DEFINITION AV805399 RAFL9 Arabidopsis thaliana cDNA clone RAFL09-42-M16 3',
mRNA sequence.

ACCESSION AV805399

VERSION AV805399.1 GI:19839384

KEYWORDS EST.

SOURCE thale cress.

ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

REFERENCE Rosidae: eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 410)
AUTHORS Seki,M., Narusaka,M., Ishida,J., Kamuya,A., Satou,M., Nakajima,M.,
Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y.
and Shinozaki,K.

TITLE Large scale analysis of Arabidopsis full-length cDNA (2002b)
JOURNAL Unpublished (2002)

COMMENT

Contact: Motoaki Seki

Plant Functional Genomics Research Group

RIKEN Genomic Sciences Center

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel: 81-298-36-4359

Fax: 81-298-36-9060

Email: meskierc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998). cDNA cleaved with BamHI and XhoI was ligated to modified lambda FLc-1 vector (Carninci et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified Bluescript vector. Please visit our web site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further details.

FEATURES

source Location/Qualifiers

1..410
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone="RAFL09-42-M16"
/dev_stage="plants at various developmental stages from germination to mature seeds"
/lab_host="DH10B"
/note="Site:1: BamHI; Site:2: SalI; subjected to dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24 hr) treatments"
BASE COUNT 129 a 94 c 74 g 113 t
ORIGIN

Query Match 20.2%; Score 338.2; DB 10; Length 410;
Best Local Similarity 98.1%; Pred. No. 5.5e-94;
Matches 354; Conservative 0; Mismatches 3; Indels 4; Gaps 1;

QY 1306 CTAGTATTTTCAGGGCTAAAGCTGTTGACGTATCTGATCAGACAAATTACTTTGACAGCTTA 1365
DB 410 CTAGTATTTTCAGGGCTAAAGCTGTTGACGTATCTGATCAGACAAATTACTTTGACAGCTTA 351
QY 1366 CTGGGATTTAACAAGATGTTGCACTGCAAAAGTTATTGGAGCCCTATGCTATATGTG 1425
DB 350 CTGGGATTTAACAAGATGTTGCACTGCAAAAGTTATTGGAGCCCTATGCTATATGTG 291
QY 1426 AGTTGCAAGAACCGGTCGTGTGCGATTGCGTGAATCGGGAGTGGAGCTCCAGATACC 1485
DB 290 AGTTGCAAGAACCGGTCGTGTGCGATTGCGTGAATCGGGAGTGGAGCTCCAGATACC 231
QY 1486 TTCGTGATATCTCTTTCTTTTAACAGGCTAAACCGTTGSCAGAGTCATCCATCGAACAT 1545
DB 230 TTCGTGATATCTCTTTCTTTTAACAGGCTAAACCGTTGSCAGAGTCATCCATCGAACAT 171
QY 1546 CAGAAACTTTGGAAGTTAAAGTTTCATTACACAGTCTATGAAAC--CTCAAAACAGCA 1601
DB 170 CAGAAACTTTGGAAGTTAAAGTTTCATTACACAGTCTATGAAACCAACCAACAGCA 111
QY 1602 CAGAGAGCTGCGTGAATATGTTGTGACTTTGTTTATGAACAATTAGCTGATTTTG 1661
DB 110 CAGAGAGCTGCGTGAATATGTTGTGACTTTGTTTATGAACAATTAGCTGATTTTG 51
QY 1662 G 1662
DB 50 G 50

RESULT 14

AW310049 607 bp mRNA linear EST 02-DEC-2001
LOCUS AW310049

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Huntville, AL 35801 For further information
call: (800)-533-4363 or contact via email: ccu@resgen.com
Seq primer: -40RP from Gldco
High quality sequence stop: 411.
Location/Qualifiers
1. 675
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl012-3211"
/clone_lib="Gm-cl012"
/tissue_type="Apical shoot tips, 9-10 day old etiolated
seedlings"
/lab_host="XL10-Gold"
/note="Vector: pBluescript II XR; Site.1: EcoRI; Site.2:
XhoI; This cDNA library was constructed from mRNA isolated
from the apical shoots of 9 to 10 day old etiolated
seedlings. The shoot tips including any emerged leaves
were harvested for mRNA isolation. The cDNA library was
prepared using the Stratagene pBluescript II XR cDNA
library construction kit. Complementary DNA was
synthesized from mRNA using a primer consisting of a poly
(dT) sequence with a XhoI restriction site. EcoRI adapters
were ligated to the blunt-ended cDNA fragments followed by
XhoI digestion. The cDNA fragments were directionally
cloned into the EcoRI-XhoI restriction site of the
pBluescript vector. The ligated cDNA fragments were
transformed into XL10-Gold host cells. This library was
constructed by Dr. Randy Shoemaker and Dr. John
Erdeljing."

BASE COUNT 193 a 125 c 170 g 187 t
ORIGIN

Query Match 19.9%; Score 333.2; DB 10; Length 675;
Best Local Similarity 70.3%; Pred. No. 2.8e-92;
Matches 475; Conservative 0; Mismatches 198; Indels 3; Gaps 2;

```
OY 625 TAGATATAGCGGACATCATGACTATGAGTAATCGAGATCCTGGAAAAATGATTG 684
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 1 TGGATATCTCCCAACATTCGGTACATTTGAGTCAATCGAGATCCAGGAAGATGCGTG 60
OY 685 CTGTAGAAGAATTTGAAAAAGTTTCAGATCAGAGACATGTTAAGACAGGAAGATAG 744
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 61 CGGTTCAAGAATTTTCAGCAAGTTTGAATTAAGAGATGCCAGAACTGGAAAG-TTG 119
OY 745 CACTGAGAAGGAAAAAGATGGGTGCACTGCTCATTTTGGCGATTTTCAGCAGCATCCT 804
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 120 CATTAAGAAGGAAAAAGATGGGTGCTATGCTCCATTTTGGCGATTTTCAGCAGCATCCT 179
OY 805 ATCCAGATCTCAAGGAGCAGCGCTGTTAGTG--TTCTTGAGATAGCAAAAAAAGAGAC 862
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 180 ATCCAGATCTTGAAGGAGAACACCTGTGAATGCCATGTGGAGCAAAAAATATGAAC 239
OY 863 CATGTCCCTCAAAAGGAACATCAGCAGGGGAGATGTTTATCCCGTTGAGCCATTTT 922
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 240 CTGTTGCCAACCTTGATCCAACTGTGGGGGAGATTGTTATCTTATAGACCACAG 299
OY 923 TGACCCCAAGGATACATGATTTCTGACGCTCATCGGGAGCTTTCAGTCAGCAAGATAC 982
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 300 TGGTTTACAGCATCAAGTACTGATGCTCAGTGGGTGCTCTCAATGAAGAAGATAC 359
OY 983 GAGTGAAGTGGTGGCATCTATCATTTGCTTGAATGATATTCAGGATTTCTTAA 1042
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 360 TAGTGAATTCGATACACACTCTATCATGCTTGTGAACGATGCTCTCGAGTCTTAA 419
OY 1043 TATGTGACTGGTGTTCGCTCGAAGGAGATACATATCCAGAGCTTGCCGTAGGACA 1102
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 420 CATGTGTACAGAGTTATTTGCTAGAAGAGGCTATTAACATTCAGAGTTAGCTAGGACA 479
```

```
OY 1103 TGCTGAACCAAGGGCATTTGACGCTTACACAGTTATACCTGCGACAGATGATCGGT 1162
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 480 TGCAGAAAGTTGAATGACTCTCGACTTACGACTGTGCTCCTCGACAGATGAGTCAAT 539
OY 1163 CAGCAAAATGGTGCAGCACTTTTACAACTGCTAGATGCTGATGCTTAC 1222
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 540 AAGCAAGTTGGTGCAGCAACTCTATTAATCTAGTAGCTACATGAGTTCGTGATATCAG 599
OY 1223 TCATTGGCATTTTCTGAAAGAACTGATGCTGATTAAGATTGGCCGTAAGCGTGTGTC 1282
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 600 CCACCTGGCATTTTCTTACCATGATGATGCTGCTGATTAAGATTGCTGCTGCTGT 659
OY 1283 TAGAAGAGATGCTCTG 1298
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 660 ACCGCGTGTATGACTG 675
```

Search completed: March 9, 2003, 21:40:21
Job time : 2579 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 9, 2003, 21:16:33 ; Search time 138 Seconds
(without alignments)
8097.319 Million cell updates/sec

Title: US-09-997-900-1

Perfect score: 1673
Sequence: 1 gctcttcagtagcaaaaa.....tgatttcggtcatttcg 1673

Scoring table:
IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 478924 seqs, 33395956 residues

Total number of hits satisfying chosen parameters: 957848

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCIT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCITUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1673	100.0	1673	10	US-09-997-900-1
2	875.6	52.3	960	10	US-09-770-445-311
3	701.8	41.9	1721	10	US-09-732-618-17
4	665.8	39.8	1861	10	US-09-732-618-23
5	593.8	35.5	1435	10	US-09-732-618-15
6	590	35.3	1813	10	US-09-732-618-9
7	537.6	32.1	1297	10	US-09-732-618-5
8	419	25.0	4895	10	US-09-997-900-3
9	243.4	14.5	611	10	US-09-732-618-1
10	201.8	12.1	724	10	US-09-732-618-21
11	197	11.8	470	10	US-09-732-618-3
12	169.4	10.1	515	10	US-09-732-618-11
13	160.8	9.6	484	10	US-09-732-618-13
14	149.4	8.9	417	10	US-09-732-618-19
15	116.8	7.0	233	10	US-09-923-618-1764
16	97.8	5.8	547	10	US-09-732-618-7
17	90	5.4	501	10	US-09-815-242-9274
18	90	5.4	501	10	US-09-815-242-9497
19	74.4	4.4	516	9	US-09-738-626-3458

C	20	64.8	3.9	640681	10	US-09-790-988-1	Sequence 1, Appl1
	21	64.6	3.9	522	10	US-09-815-242-9920	Sequence 9920, Ap
	22	62.6	3.7	489	10	US-09-815-242-3987	Sequence 3987, Ap
	23	62.6	3.7	522	10	US-09-815-242-7645	Sequence 7645, Ap
	24	56.8	3.4	492	10	US-09-815-242-7946	Sequence 7946, Ap
	25	55.4	3.3	492	10	US-09-761-782-1	Sequence 1, Appl1
	26	55.4	3.3	492	10	US-09-815-242-5926	Sequence 5926, Ap
	27	54.6	3.3	492	10	US-09-815-242-7159	Sequence 7159, Ap
	28	51.4	3.1	2000	9	US-09-938-842A-5198	Sequence 5198, Ap
C	29	38.4	2.3	8921	10	US-09-070-927A-68	Sequence 68, Appl
	30	36.6	2.2	273	10	US-09-878-574-14178	Sequence 14178, A
	31	36.2	2.2	2088	10	US-09-809-790-3	Sequence 3, Appl1
	32	36.2	2.2	2088	10	US-09-809-617-3	Sequence 3, Appl1
C	33	35.8	2.1	1754	10	US-09-998-598-359	Sequence 39, Appl
C	34	35.8	2.1	2443	9	US-09-802-640-39	Sequence 39, Appl
	35	35.2	2.1	3294	10	US-09-864-864-268	Sequence 268, Ap
	36	35.2	2.1	3344	10	US-09-964-824A-550	Sequence 550, Ap
	37	35.2	2.1	3344	10	US-09-954-456-2146	Sequence 2146, Ap
	38	35.2	2.1	3344	10	US-09-880-107-2260	Sequence 2260, Ap
C	39	35	2.1	118	10	US-09-815-242-1077	Sequence 1077, Ap
	40	35	2.1	2780	10	US-09-841-786-10	Sequence 10, Appl
	41	35	2.1	9726	10	US-09-841-786-8	Sequence 8, Appl
	42	35	2.1	11130	10	US-09-841-786-15	Sequence 15, Appl
C	43	34.4	2.1	3660	10	US-09-070-927A-112	Sequence 112, Ap
	44	33.8	2.0	255	10	US-09-923-876-1370	Sequence 1370, Ap
	45	33.8	2.0	4140	10	US-09-764-877-3620	Sequence 3620, Ap

ALIGNMENTS

RESULT 1

US-09-997-900-1

Sequence 1, Application US/09997900

Patent No. US20020053098A1

GENERAL INFORMATION:

APPLICANT: Kakufuda, Genichi

APPLICANT: Costello, Colleen

APPLICANT: Sun, Ming

APPLICANT: Hu, Weiming

TITLE OF INVENTION: Genes and Vectors for Conferring Herbicide Resistance

FILE REFERENCE: 043753/241148 (5849-20A)

CURRENT APPLICATION NUMBER: US/09/997,900

CURRENT FILING DATE: 2001-11-30

PRIOR APPLICATION NUMBER: 60/106,239

PRIOR FILING DATE: 1998-10-29

PRIOR APPLICATION NUMBER: 09/426,568

PRIOR FILING DATE: 1999-10-22

NUMBER OF SEQ ID NOS: 11

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 1

LENGTH: 1673

TYPE: DNA

ORGANISM: Arabidopsis sp.

FEATURE:

NAME/KEY: CDS

LOCATION: (42)..(1514)

OTHER INFORMATION: Mature Peptide

US-09-997-900-1

Query Match 100.0%; Score 1673; DB 10; Length 1673;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 1673; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCTCTTCAGTAGCAAAAACCTTCGCTCTGCTCAATGGCGCCATTCTGTAA 60
DB 1 GCTCTTCAGTAGCAAAAACCTTCGCTCTGCTCTGCTCAATGGCGCCATTCTGTAA 60
QY 61 GCTCTTCAGTAGCAAAAACCTTCGCTCTGCTCTGCTCAATGGCGCCATTCTGTAA 120
DB 61 GCTCTTCAGTAGCAAAAACCTTCGCTCTGCTCTGCTCTGCTCAATGGCGCCATTCTGTAA 120

QY	121	TATCTGAGCGCGTATGCTTCCCGCGCAAGTTTCATATCTCCGATATATCTTCC	180
Db	121	TATCTGAGCGCGTATGCTTCCCGCGCAAGTTTCATATCTCCGATATATCTTCC	180
QY	181	ACCGTGGCGATGAATGGGTAAAGAGATGGAAGGATTCGTTAAGACGCTCGATGGAGA	240
Db	181	ACCGTGGCGATGAATGGGTAAAGAGATGGAAGGATTCGTTAAGACGCTCGATGGAGA	240
QY	241	TCCTGATGCGCTCTTTCYCCGAAGCTTCATCTGCGACTCCAAATTCGAAGGTGAGAGC	300
Db	241	TCCTGATGCGCTCTTTCYCCGAAGCTTCATCTGCGACTCCAAATTCGAAGGTGAGAGC	300
QY	301	ACACAAATTCAGTATTTGTTGGAGACGAAAGCGGAATGATATATAGATTGCGAGAGCT	360
Db	301	ACACAAATTCAGTATTTGTTGGAGACGAAAGCGGAATGATATATAGATTGCGAGAGCT	360
QY	361	TTGCAGAGAGAGATACAAATATTGAGAGTCTTGCTGTGTAACAGACAGAGCTC	420
Db	361	TTGCAGAGAGAGATACAAATATTGAGAGTCTTGCTGTGTAACAGACAGAGCTC	420
QY	421	TATTCACCATAGTTGTCTGGAAGTGAAGGGTACTTCAGCAGGTCATCGCAACTCC	480
Db	421	TATTCACCATAGTTGTCTGGAAGTGAAGGGTACTTCAGCAGGTCATCGCAACTCC	480
QY	481	AGAAGCTCTGTAATGTCTTAAAGGTTGAAGATATCTCAAGTACGCCGAATGAGCGTG	540
Db	481	AGAAGCTCTGTAATGTCTTAAAGGTTGAAGATATCTCAAGTACGCCGAATGAGCGTG	540
QY	541	AGCTGATGCTGTGTAAGTGAATGCATCCAGATCCAGAGTCCAGGGCAGAGATCATGTGGTAG	600
Db	541	AGCTGATGCTGTGTAAGTGAATGCATCCAGAGTCCAGAGTCCAGGGCAGAGATCATGTGGTAG	600
QY	601	TTGACACATTCAGACGACAAGTGTAGATATAGCGGACATGCATTTGACTATCGAGTAA	660
Db	601	TTGACACATTCAGACGACAAGTGTAGATATAGCGGACATGCATTTGACTATCGAGTAA	660
QY	661	CTGGAAGTCTCGGAAAAATATATTGCTGTGTAAGAAATTTGAAAAAGTTTCAGATCAGAG	720
Db	661	CTGGAAGTCTCGGAAAAATATATTGCTGTGTAAGAAATTTGAAAAAGTTTCAGATCAGAG	720
QY	721	AGATTGTAAGACGAGAAATAGCACTGAGAAAGGAAAGATGGGTGCACACTGCTCAT	780
Db	721	AGATTGTAAGACGAGAAATAGCACTGAGAAAGGAAAGATGGGTGCACACTGCTCAT	780
QY	781	TTTGGCGATTTTCACGAGCATCTCATCAGATCTCAAGAGCAGACGCGCTGTATGATGTC	840
Db	781	TTTGGCGATTTTCACGAGCATCTCATCAGATCTCAAGAGCAGACGCGCTGTATGATGTC	840
QY	841	TTTCAAGTACGAAAAAGAGCCATTGTGCTCAAAAGCAAAATCAGAGGGGAGATG	900
Db	841	TTTCAAGTACGAAAAAGAGCCATTGTGCTCAAAAGCAAAATCAGAGGGGAGATG	900
QY	901	TTTATCCGTTGACCCATTTTGTGACCCCAAGGTACATGTATTTCTGACGCTCAGCTGGG	960
Db	901	TTTATCCGTTGACCCATTTTGTGACCCCAAGGTACATGTATTTCTGACGCTCAGCTGGG	960
QY	961	GACTTCTACGAGAGATACGATGAGTGCATCGGTCGATCTATCATCTTCTGTAA	1020
Db	961	GACTTCTACGAGAGATACGATGAGTGCATCGGTCGATCTATCATCTTCTGTAA	1020
QY	1021	ATGATATTCACGAGCTCTTAAATATTTGTGACTGTGTTTTCGCTCGAAGGGGATACAATA	1080
Db	1021	ATGATATTCACGAGCTCTTAAATATTTGTGACTGTGTTTTCGCTCGAAGGGGATACAATA	1080
QY	1081	TTCCAGAGCTTGGCCGTGAGCATGCTGAAACCAAGGGCATTTTCACGCAATTCACACAGTTA	1140
Db	1081	TTCCAGAGCTTGGCCGTGAGCATGCTGAAACCAAGGGCATTTTCACGCAATTCACACAGTTA	1140
QY	1141	TATCTGCAACAGATGAATGCGTACAGCAAAATGGTGAGCAACTTTACAAACGTCGATG	1200
Db	1141	TATCTGCAACAGATGAATGCGTACAGCAAAATGGTGAGCAACTTTACAAACGTCGATG	1200
QY	1201	TGCATGAGGTCATGCTTACTCATTTTCCCATTTTTCGAAAGAGAACTGATGCTGATTA	1260

Db	1201	TCGATGAGGTCACGATCTTACTTCATTTGGCAATTTCTGAAGAAACGATGATCGATTA	1260
Qy	1261	AGATTGCCGTGAACGCGTGTGTAGAAAGATGTCCTGAGACATTCATGTATTTTCACGG	1320
Db	1261	AGATTGCCGTGAACGCGTGTGTAGAAAGATGTCCTGAGACATTCATGTATTTTCACGG	1320
Qy	1321	CTAAAGCTGTTGACATATCTGATCACAACAATTACTTGCAGCTTACTGGGATCTAGACA	1380
Db	1321	CTAAAGCTGTTGACATATCTGATCACAACAATTACTTGCAGCTTACTGGGATCTAGACA	1380
Qy	1381	AGATGCTTGCACATGCAAAAGTTATTTGAGCCCTATGCTATATGTAGAGTTGCAGAACCG	1440
Db	1381	AGATGCTTGCACATGCAAAAGTTATTTGAGCCCTATGCTATATGTAGAGTTGCAGAACCG	1440
Qy	1441	GTCGCTGGCGCATTTGGCGCGATTCGGAGTGGAGTCCCAAGTACTTGGTGATATCTCT	1500
Db	1441	GTCGCTGGCGCATTTGGCGCGATTCGGAGTGGAGTCCCAAGTACTTGGTGATATCTCT	1500
Qy	1501	TTCTTTTAAACAGGCTAAACCGTTGACAGAGTGCATCCATGGAACATCAGAAACTTTGGAG	1560
Db	1501	TTCTTTTAAACAGGCTAAACCGTTGACAGAGTGCATCCATGGAACATCAGAAACTTTGGAG	1560
Qy	1561	GTAAAGCTTTCATTACACAGCTATATGAACCTCAAGACAGACAGAGAGACTGCGTGGATA	1620
Db	1561	GTAAAGCTTTCATTACACAGCTATATGAACCTCAAGACAGACAGAGAGACTGCGTGGATA	1620
Qy	1621	TATGTTTGTGACTTGTATTATGAACAATTTAGCTGATTTTGGGCTTCATTTG	1673
Db	1621	TATGTTTGTGACTTGTATTATGAACAATTTAGCTGATTTTGGGCTTCATTTG	1673

```

US-09-770-445-311/c
; Sequence 311, Application US/09770445
; Patent No. US200203281A1
; GENERAL INFORMATION:
; APPLICANT: Goriach, Jörn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Ralnes, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Mathew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krickler, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; TITLE OF INVENTION: thaliana
; FILE REFERENCE: 2023US (PARA-012PRV)
; CURRENT APPLICATION NUMBER: US/09/770,445
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/178,472
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 311
; LENGTH: 960
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-445-311

Query Match      52.3%   Score 875.6;   DB 10;   Length 960;
Best Local Similarity  98.7%   Pred. No. 2,4e-262;

```


RESULT 5
 US-09-732-618-15
 Sequence 15, Application US/09732618
 Patent No. US2001004939A1
 GENERAL INFORMATION:
 APPLICANT: Abell, Lynn
 APPLICANT: Falco, Carl
 APPLICANT: Famodu, Omolayo O.
 TITLE OF INVENTION: Small Subunit of Plant Acetolactate Synthase
 FILE REFERENCE: B61435 US NA
 CURRENT APPLICATION NUMBER: US/09/732,618
 PRIOR FILING DATE: 2000-12-08
 PRIOR APPLICATION NUMBER: 60/174,437
 NUMBER OF SEQ ID NOS: 43
 SOFTWARE: Microsoft Office 97
 SEQ ID NO 15
 LENGTH: 1435
 TYPE: DNA
 ORGANISM: *Oryza sativa*
 US-09-732-618-15

Query Match	35.5%;	Score 593.8;	DB 10;	Length 1435;
Best Local Similarity	71.3%;	Pred. No. 2e-174;		
Matches 813;	Conservative	0;	Mismatches 322;	Indels 6;
				Gaps 2;

OY	394	CTTGTGCTGAAAGAGACAAAGGCTCTATTACCACTAATTGTCTGTGGAACTGAAGGG	453
Db	19	CCGGGGGGCTCAACAAAGACAAAGGCCATGTCTCAACATTTGCTCTCCGGCACGGACGGG	78
OY	454	TACTTTCAGCAGGTCATCAGACCACTCCAGAACGCTCTTAATGTTCTTAAAGTTGAAGATA	513
Db	79	TGCTCAACCAAGTCAATCGAGACACTCAACAAGCTTGTCAAGCTCTTGAATGTGGAAAGTC	138
OY	514	TTCTAAGTAGCCGCGCAAGTGGAGCGTAGCTGATGCTGTAAAGTGAATGCACATCCAG	573
Db	139	TATCTAAGGAGGCCACAGGTTGAAGAGAGAGCTGATGCTTATTAATAATTTAATGTTGAACAG	198
OY	574	AATCCAGGCGAGANTCATATGCGCTAGTTGACACATTCCAGACCAAGAGTTGATAGATAG	633
Db	199	ATACAGCGTCTGAGCTCATATGTTTATGTTGATATATTTTCCGACGCAAGAGTGTGATATTT	258
OY	634	CGGAACTATGCATTCATATGAGGTAACTGAGATCCCTGGAATAATGATTCGTGTAGAAA	693
Db	259	CGGAGAACACCCCTTACCATCGAGGTAACTGAGATCCTTGCGCAAAATTTGTTGCTGTGCAAA	318
OY	694	GAAATTTGAAAAAGTTTCAGATCAGAGAGATGTGAAGCAGAGAAAGATAGCATGAGAA	753
Db	319	GGAACCTCAGCAAAATTTGGATTAAGAAATTTGTGAACGGGAAAAATTCCTTTGAGAC	378
OY	754	GGGAAAAAGATGGGTCGAACTGTCCTCAATTTTGGCGATTTTCACAGACATCTTATCCAGATC	813
Db	379	GTGAAAAAATTTGGAGCAACGACCCCGCTTCCTGGGGAATTTCTCTGCTTCTTACCAAGATC	438
OY	814	TCAAGGAGACAAAGCCGCTAGTGTAGTCTTCTTCGA - - GTAGCAAAAAAGAGACCATTTGCC	870
Db	439	TCTATAGAGGCACTTGGCCCAAAATTTCTTCTTACTTCTGTAAATAGACAGCATATGGAA	498
OY	871	CTCAAAAAGGAACATC - - AGCAGGGGAGAGTGTATCCGCTGAGCCATTTTTTGACC	927
Db	499	GTTTGTATCAACCAATCCCAATGCTGGGGGCGATGTATCTGTGMACTTATAGACGTT	558
OY	928	CCAAGGTACATCTGATTTCTGAGGCTTCACAGGCGGACCTTCTCACTAGCAAGATACAGATG	987
Db	559	CATCCATGAACCAAGTACTTGTATGCTCTACCTGGGGCGCTCCTTGATGAAGATTAACGTG	618
OY	988	GACTACGCTGCATACATCTATCATATCTCTGTAATGATATTCAGAGAGTCTTAATATTTG	1047
Db	619	GACTTCGATCAATACATCTATCCATCTCTTGTCAATATATTTGCCCCGTGTTCTTCACATTTG	678
OY	1048	TGACTGTGTTTTCGCTCGAAGGGAGATACAAATTCACAGAGCTTGGCCGTAGGACATGCTG	1107

Dd	679	TTACAGGGGCTTTCGTCGCAGAGGCTACAATATACAGAGTCCTTGCTAGGCCACACTG	738
Oy	1108	AAACCAAGGCGATTTCACGCGATTACAACAAGTTATACCTGCAACAGANTGCGGTACGA	1167
Dd	739	AAAAGTCAGGCGCTTCGCGTATATACAACAGTTGCTCCTCGAACAAGATGAATTCATTGAGA	798
Oy	1168	AATTGGTAGCAACACTTATAAACCTGTAAGTGTGATAGAGTCCATGATCTTACTATT	1227
Dd	799	AGTTAGTTCAGCACCTTAACAACACTGGTATGATGTGATGATGATGAAGTTCAAATATTAATCTACT	858
Oy	1228	TGCCATTTTCTGAAGAAGAACATGATGCTGATTAAGATTGCCCTGAACGCTGCTGTAGAA	1287
Dd	859	TGCCTTTTCGTGAAGAAGAACCTTATGCTTATATCAAGSTTCTGTGAACACTGGTCTCGGA	918
Oy	1288	GAGATGTCCTGAGACATTGCTAGTATTTTTCAGGCGCTAAACCTGTTAGCATATCTGATACA	1347
Dd	919	GAGACATACAGATATATGGTCGAAGATCTTCGCGGCAAAAACTGTATGATTTCTGTGATACA	978
Oy	1348	CATTATCTTTGACACTTACTGCGGAGATCTAGACAAAGTGGTGCACATGCAAAAGTTATATGG	1407
Dd	979	CTGTACCTTACACCTTACTACGCGGAGATCTGCACAAAGATGTTTCATTACAAACGCTTTGG	1038
Oy	1408	AGCCCTATGTAATATGTGAGGTTCAGAAACGCGTCGTGTGSCAATYGCCTGTGANTCG	1467
Dd	1039	AGCCTTAATGGCATCTGTGAAGTCCGCCAAGAACAGGGGAGATGGCGCTGGATCCGCGCATATCG	1098
Oy	1468	GAGTGAAGTCCCAATACTACTCTGAGATATACCCTTTCTTTTAAACAGGCTAAACCCGTTAG	1537
Dd	1099	GTCGATATCCAAAGTACCTTCGTGGCTACTCTTCCTGTTGTAATCCCAAGGCTTTGTGAG	1158
Oy	1528	A	1528
Dd	1159	A	1159

```

: RESULT 6
: US-09-732-618-9
: Sequence 9, Application US/09732618
: Patent No. US20010044939A1
: GENERAL INFORMATION:
: APPLICANT: Abell, Lynn
: APPLICANT: Falco, Carl
: APPLICANT: Farnodu, Onolayo O.
: TITLE OF INVENTION: Small Subunit of Plant Acetylactate Synthase
: FILE REFERENCE: BB1435 US NA
: CURRENT APPLICATION NUMBER: US/09/732,618
: CURRENT FILING DATE: 2000-12-08
: PRIOR APPLICATION NUMBER: 60/174,437
: PRIOR FILING DATE: 2000-01-04
: NUMBER OF SEQ ID NOS: 43
: SOFTWARE: Microsoft Office 97
: SEQ ID NO 9
: LENGTH: 1813
: TYPE: DNA
: ORGANISM: Zea mays
: US-09-732-618-9

```

Query Match	35.3%	Score 590;	DB 10;	Length 1813;
Best Local Similarity	68.2%;	Pred. No. 3.6e-173;		
Matches 834;	Conservative 0;	Mismatches 385;	Indels 3;	Gaps 1;

QY	285	TCGAAAGTGAGGACGACACAAATTTCTAGTATTTCTGTGCACGACGAAAGCGCAATCATTAAT	3444
Db	223	TCGGTTATTAAGAACGCTCACACGCTATTCACATTTTGTGTGTATGAAGATGGGATGATCAAT	2822
QY	345	AGGATTGCAGCAGCTGTTTTCGCAAGCAGAGGATACAAATATTGAGAGTCTTGCTGTGGTCTG	4040
Db	283	CGAATTCGTGGGGTTTTCGTAGAAAGAGATTAACATCGAGTCATTGGCTGTGGCGTTG	3422
QY	405	AACACAGACAGGCTCTATTACCATATTTCTGTGTGAACTGAAAGAGTACTTTCACACAG	4646
Db	343	AACAAAGGTTAAAGCTTTTATTACATATATAGTGTGCAGGAAACGCAAAATTTTAAACAG	4020

OY	1180	AACTTTTCAAACTGAGTAGATGTCACATGAGAGTCGATCGTACATTTGCAATTTCTG	1239
Db	723	ACGTTTCAAAAGCTATTATTGATGTGCATGAAGTTCAATGACATTAACCACTACCTTTTGCTG	782
OY	1240	AAAGAGAACTGATGCTGATTAAGATTTAGCTGCGCGTGAACGCTGCTGCTAGAGAGATGCTCGG	1299
Db	783	AAAGGAGAGCTGATGCTATTATTAAGTTTCTGTAAACACTGCTGCTGAGGAGGAAATTTCTAG	842
OY	1300	ACATTTGCTAGTATTTTTCAGGGCTAAAGCTGTTGACGTATCTGATCACAAATTACTTTGC	1359
Db	843	ATATTGCTGAATCTTTCCGAGCAAAACGTATTGAGTTTCTGACATPACAGTAACCCCTTC	902
OY	1360	AGCTTACTGGGGACTTAGACAGATGTTGCTCACTGCAAAAGTTATTGAGACCCCTATGCTA	1419
Db	903	AGCTTACTGGAGATCTTTCGACAAAGATGTTGCTCACTCAAAAGGTTATTAGAACCCATATGGCA	962
OY	1420	TATTTGAGGTTGCAAAACCGGTCTGTGTGGCATTTGGTCTGTAAATCGSGAGATGGACTTCA	1479
Db	963	TTCTGCGAGGTGCGCAAAACGAGCAAGATGGCACTGGTGTGTAATGGAAGGTGCAGCTTCCA	1022
OY	1480	AGTACTTTCGTGATACTCCTTTCTCTTT	1507
Db	1023	AGTACCTCGCGGTACTCTCTTCCATT	1050

```

RESULT 8
US-09-997-900-3
; Sequence 3, Application US/09997900
; Patent No. US20020053098A1
; GENERAL INFORMATION:
; APPLICANT: Kakefuda, Genichi
; APPLICANT: Costello, Colleen
; APPLICANT: Sun, Ming
; APPLICANT: Hu, Weiming
; TITLE OF INVENTION: Genes and Vectors for Confering Herbicide Resistance
; TITLE OF INVENTION: in Plants
; FILE REFERENCE: 043753/241148 (5849-20A)
; CURRENT APPLICATION NUMBER: US/09/997,900
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: 60/106,239
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: 09/426,568
; PRIOR FILING DATE: 1999-10-22
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 4895
; TYPE: DNA
; ORGANISM: Arabidopsis sp.
; FEATURE:
; NAME/KEY: Promoter
; LOCATION: (1)..(757)
; OTHER INFORMATION: Promoter Region
; NAME/KEY: misc_feature
; LOCATION: (717)
; OTHER INFORMATION: Transcriptional Startling Point
; NAME/KEY: misc_signal
; LOCATION: (758)..(760)
; OTHER INFORMATION: Start Codon
; NAME/KEY: misc_signal
; LOCATION: (4737)..(4739)
; OTHER INFORMATION: Stop Codon
; OTHER INFORMATION: n at position 694 can be a, c, g, or t
US-09-997-900-3

```

Query Match	25.0%	Score 419	DB 109	Length 4895
Best Local Similarity	86.7%	Pred. NO. 1.3e-119		
Matches	507	Conservative	0	Mismatches 0; Indels 78; Gaps 1
QY	1	GTCCTTCACGATGACAAAACCTTCGGCTTCGTCTCGTCAATGGCGGCACATTTCTGAA	60	
Db	717	GTCCTTCACGATGACAAAACCTTCGGCTTCGTCTCGTCAATGGCGGCACATTTCTGAA	776	

QY	61	GTGTTACATCATCTGCTGCTTGAGATCGCATGTTCCGATCTCTCTGCTTG	120
Db	777	TTTTTTTACCATCTATCTGCTGCTTGAGATCGCATGTTCCGATCTCTCTGCTTG	836
QY	121	TATCCTCGACGCGTATGCTGTTCCGCGCAAGATTTCATCTCCGGTATATCTCGC	180
Db	837	TATCCTCGACGCGGTATGCTGTTCCGCGCAAGATTTCATCTCTCCGGTATATCTCGC	896
QY	181	ACCGTGCAGATGAATATGGTAAAGAGAATGGAAGATTTCTTGAAGCGCTCATGGGA	240
Db	897	ACCGTGCAGATGAATATGGTAAAGAGAATGGAAGATTTCTTGAAGCGCTCATGGGA	956
QY	241	TCTGTGATGCGTCTTCTCCGAAGCTTTCATCTGCGACTCCAAATGCA-----	289
Db	957	TCTGTGATGCGTCTTCTCCGAAGCTTTCATCTGCGACTCCAAATGCAAGCGACTGTGA	1016
QY	290	-----	289
Db	1017	TAAATATTTGCTTAAAGTCGTTTCTTTTGCGCTTTGCTTGTGATTTCTTTGTGCATTA	1076
QY	290	-----GGTGAGGAACACACAATTTCTAGTATTTTGTGAGACCAAGCGAATGATTA	342
Db	1077	AAATCAGGCGTGAAGACACACAATTTCTAGTATTTTGTGAGACCAAGCGAATGATTA	1136
QY	343	ATAGGATTTGCAGGAGTGTGTTGCAAGGAGAAGATACAAATTGAGAGTCTTGCTGTTGCTC	402
Db	1137	ATAGGATTTGCAGGAGTGTGTTGCAAGGAGAAGATACAAATTGAGAGTCTTGCTGTTGCTC	1196
QY	403	TGAACAGACAGACAGGCTTATTCACATATGTTGCTGTGGAACCTAAGAGGTACTTCAGC	462
Db	1197	TGAACAGACAGACAGGCTTATTCACATATGTTGCTGTGGAACCTAAGAGGTACTTCAGC	1256
QY	463	AGGTCATCGAGCAACCTCAGAAAGCTGTTAATGTTCTAAGGTTG	507
Db	1257	AGGTCATCGAGCAACCTCAGAAAGCTGTTAATGTTCTAAGGTTG	1301

```

RESULT 9
US-09-732-618-1
; Sequence 1, Application US/09732618
; Patent No. US20010044939A1
; GENERAL INFORMATION:
; APPLICANT: Abell, Lynn
; APPLICANT: Falco, Carl
; APPLICANT: Famodu, Omolayo O.
; TITLE OF INVENTION: Small Subunit of Plant Acetolactate Synthase
; FILE REFERENCE: BBI435 US NA
; CURRENT APPLICATION NUMBER: US/09/732,618
; CURRENT FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/174,437
; PRIOR FILING DATE: 2000-01-04
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 1
; LENGTH: 611
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (61)
; NAME/KEY: unsure
; LOCATION: (328)
; NAME/KEY: unsure
; LOCATION: (331)
; NAME/KEY: unsure
; LOCATION: (357)
; NAME/KEY: unsure
; LOCATION: (361)
; NAME/KEY: unsure
; LOCATION: (364)
US-09-732-618-1

```



```

; LOCATION: (452)
; NAME/KEY: unsure
; LOCATION: (454)
US-09-732-618-3

```

Query Match	11.8%;	Score 197;	DB 10;	Length 470;
Best Local Similarity	64.3%;	Pred. No. 4.5e-51;		
Matches 297;	Conservative 0;	Mismatches 162;	Indels 3;	Gaps 1;

QY	460	AGCAGGTCATGAGACCACTCAGAAAGCTCGCTTTATGTTCTTAAAGGTGAAGATATGCA	519
Db	1	ACCAAGTCATTGAGCACTCAATTAAGTCTGTCACAGCTTCACTAGTGTGGAAGATCTATCTA	60
QY	520	GTCAGCCGCAAGTGGAGCGTAGCTGATGCTTGTAAAAGTAAATGCACATCCAGAAATCCA	579
Db	61	AGAAACCTCAGTGTGAAAGAGAGCTGATGCTTTAAAGCTAAACGNNAGAACCTGATCAGC	120
QY	580	GGGAGAGATCATGTGGCTAGTGTACACATTCCAGAGCAAGAAGTTGTAGATATATAGCGAAC	639
Db	121	GCCCTGAGGNCATGTTTATGTTGATCATTCTTCAGAGCAAAAAGTNGTGAATAMNCAGAG	180
QY	640	ATGCATTGACTATGAGGTAACTGGAATCCTGGAAAAATGATTCGCTGTGAAGAAATTT	699
Db	181	AAACACTTACCANNAGAGNMGCTGGAATCTCTGGCAAAATTCGCTGCAGTCAAGAGATC	240
QY	700	TGAAAAAGTTTCAGATCAGAGAGATTGTAAGACAGGAAAAATGACATGAGAGGGA	759
Db	241	TAAAGAAATTCGCGATCAAGAAATTTGCAGAGACAGGAAAAATTCGTTGAGACGNAAA	300
QY	760	AGATGGGGCAACGTCGCCATTTTGGGGATTTTTCAGAGCATCTTANCCAGATCTCAAG	819
Db	301	AGATTGGNCAACAGCCCGTTCTGGGGATTTTCTGCTGCTCTTATACACAGACCTTAAAG	360
QY	820	AGCAAGCGCCCTGTTAAGTGTCTTTCGAAAGTAGCAAAAAAGAGAGCCATTCCTCTCAAAAG	879
Db	361	AGGCATTACCAAAAAACCGCTTCACTGTAATAAATAAGACAGTGAATAGCGAGTTTGTTC	420
QY	880	AAACATC---AGCAGGGGAGATGTTTATCCCGTTGAGCAT	918
Db	421	GACCATCCAAATGCTGGGGGTGATGTTTAAATCCNGAGAACTTT	462

RESULT 12
 US-09-732-618-11
 Sequence 11, Application US/09732618
 Patent No. US20010049439A1
 GENERAL INFORMATION:
 APPLICANT: Abell, Lynn
 APPLICANT: Falco, Carl
 APPLICANT: Farnodu, Omolayo O.
 TITLE OF INVENTION: Small Subunit of Plant Acetolactate Synthase
 FILE REFERENCE: B91435 US NA
 CURRENT APPLICATION NUMBER: US/09/732,618
 PRIOR FILING DATE: 2000-12-08
 PRIOR APPLICATION NUMBER: 60/174,437
 NUMBER OF SEQ. ID NOS: 43
 SOFTWARE: Microsoft Office 97
 SEQ ID NO 11
 LENGTH: 515
 TYPE: DNA
 ORGANISM: *Oryza sativa*
 FEATURE:
 NAME/KEY: unsure
 LOCATION: (462)
 NAME/KEY: unsure
 LOCATION: (481)
 US-09-732-618-11

Query Match	10.1%;	Score 169.4;	DB 10;	Length 515;
Best Local Similarity	65.2%;	Pred. No. 1.9e-42;		
Matches 311;	Conservative 0;	Mismatches 158;	Indels 8;	Gaps 4

OY	568	ATTCAGAAATCCAGGCGAGAGATATATGTGGCTAGTTGACACAACTTCAGACCAAGATTTGAG	627
	11		11
Db	11	AACCAAGTCAGCGCTCGTAGGTCATGGTTTATGTTGATATTTTTCGAGCGGAAGTGTGG	70
OY	628	ATTATAGCGGAACATGCATTTGACTATGAGGTAACTGGAGATCCTGGAAAAATGATGCTG	687
	71		71
Db	71	ATTATTTGGGAGAACCCCTTCATCATGAGGTAACTGAGATTCCTGGCAAAATTTGTGGTG	130
OY	688	TAGAAGAAATTTGAAAAAGTTTCAGATCAGAGATTTGTAAAGCAGAGAAATATGAC	747
	131		131
Db	131	TGCAAGAGAACCTCAGCAAAATTTGGGATTAAGAAATTTGTGAACGGGAAAAATTTGCTT	190
OY	748	TGACAGAGGAAAAAGATGGGGTGCAACTGTCACATTTTGGCGATTTACACACATCCTATC	807
Db	191	TGACACTGTAAAAAATTTGAGCACTGCCCCGCTTCTGGGGATTTTCTCTGCTTTCAC	250
OY	808	CAGATCTCAGAGGACAGACGCGCTGTAGTGTCTTCGAA--GTAGCAAAAAAGAGACCA	864
	251		251
Db	251	CAGATCTCATAGAGGCGATTTGCCCAAAATTTCTCTTACTCTTAATTAAGACAGTCA	310
OY	865	TTTGCCCTCAAAAAGAAACATC--AGCAAGGGGGAGATGTTATCCGCTTGAGCATTTT	921
	311		311
Db	311	ATTGAGAGTTTGTATCAACCAATCCAAATGCTGGGGGCGATGTCATCTTGTGAACTTATG	370
OY	922	TTGACCCCAAGGTACATGATTTCTGGAGGCTCAGTGGGAGATTCCTCATACGNAAGATA	981
	371		371
Db	371	AGGGTTTCATCATGAACCAAGTACTGATGTCACATGGGGGCTCTTGATGTAAAGT-	429
OY	982	CGAGTGGACATACGGTCCCATACTCTATCATTTGCTTGTAAATATATTTCCAGAGATTC	1038
	430		430
Db	430	CAAGTGGACTTCGATCACTACATCTCAACA--TCCATGTCACTAATTTGGCCCTGGNGTTC	485

```

US-09-732-618-13
; RESULT 13
; Sequence 13, Application US/09732618
; Patent No. US20010044939A1
; GENERAL INFORMATION:
; APPLICANT: Abell, Lynn
; APPLICANT: Falco, Carl
; APPLICANT: Famodu, Omolayo O.
; TITLE OF INVENTION: Small Subunit of Plant Acetolactate Synthase
; FILE REFERENCE: Bbl435 US NA
; CURRENT APPLICATION NUMBER: US/09/732,618
; CURRENT FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/174,437
; PRIOR FILING DATE: 2000-01-04
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 13
; LENGTH: 484
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (350)
; NAME/KEY: unsure
; LOCATION: (419)
; NAME/KEY: unsure
; LOCATION: (424)
; NAME/KEY: unsure
; LOCATION: (436)
; NAME/KEY: unsure
; LOCATION: (440)
; NAME/KEY: unsure
; LOCATION: (463)
; NAME/KEY: unsure
; LOCATION: (476)
US-09-732-618-13

```

Query Match	9.68;	Score	160.8;	DB	10;	Length	484;
Best Local Similarity	66.33;	Pred. No.	8.8e-40;				
Matches	246;	Conservative	0;	Mismatches	133;	Indels	2;
						Gaps	1;

```

Qy 394 CTGTTGCTGTAACAGACAAAGGCTCTATTCACCAATAGTGTCTGTGGAACCTGAAGGG 453
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12 CCGTGGGGCTCAACAGACAAAGGCCATGTTCACCATGTGCGCTCGGCGACGACAGGG 71
Qy 454 TACTTCAGAGGTCATCGACCACTCCAGAGCTGCTTATGTCTTAAGGTTGAAGATA 513
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 72 TGGTCAACCAAGTCATCGAGAGCTCAGACAGCTGCTCAACGCTTGAAGTGTGGAGATC 131
Qy 514 TCTCAAGTGGCCGCAAGTGGAGCCGTGAGCTGATGCTGTAAAGATGATGCACATCCAG 573
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132 TATCTAAGAGAGCCACAGTTGGAAGAGAGCTGATGCTTATAAAAATTAATGTTGAACCA 191
Qy 574 AATCCAGGGCAGAGATCATGTGGCTAGTTGACACATTCAGACAGACAGATTTAGATATAG 633
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 192 ATCAGCGCTCTGAGTCATGTGTTTAGTATATATTTCCGAGGAAAGTTGTTGATATTT 251
Qy 634 CGGACATCATTTGACTATTCAGAGTAACTGG--AGATCCTGGAATAATGTTGCTGTAA 691
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 252 CGGAGAACACCCCTTACCATTCAGAGTAACTGGAGATCTGGGGCAAAATTTGCTGTGCA 311
Qy 692 AAGAATTTGAAAAAGTTTCAGATCAGAGAGATTTGAAGACAGAGAAAGATGCACTGAG 751
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 312 AAGGAACCTCAGCAATTTGGCAAAAAAGAAATTTGTANAAGCGGCAAAATTTGCTTGA 371
Qy 752 AAGGAAAAAGA 762
    ||| ||| |||
Db 372 GACGTGAAAAA 382

```

RESULT 14

```

US-09-732-618-19
; Sequence 19, Application US/09732618
; Patent No. US20010044939A1
; GENERAL INFORMATION:
; APPLICANT: Abell, Lynn
; APPLICANT: Falco, Carl
; APPLICANT: Famodu, Omolayo O.
; TITLE OF INVENTION: Small subunit of Plant Acetolactate Synthase
; FILE REFERENCE: B81435 US NA
; CURRENT APPLICATION NUMBER: US/09/732,618
; CURRENT FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/174,437
; PRIOR FILING DATE: 2000-01-04
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 19
; LENGTH: 417
; TYPE: DNA
; ORGANISM: Trifolium aestivum
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (367)
; NAME/KEY: unsure
; LOCATION: (381)
US-09-732-618-19

```

```

Query Match      8.9%: Score 149.4; DB 10; Length 417;
Best Local Similarity 70.8%; Pred. No. 2.9e-36;
Matches 226; Conservative 0; Mismatches 91; Indels 2; Gaps 2;

```

```

Qy 363 GCAAGAGAGAGATACAATATTGAGACTCTTGCTGTGCTGMAAGAGACAGAGCTCTA 422
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 GCGAGAGAGAGGCTACAACATCGAGTCGCTCGCGCTTAACAAGAGCAAGGCCCTC 60
Qy 423 TTACCATATGTTGTCTGTGGAAGTGAAGGTAATTCACAGAGTCAATCGACAACTCCAG 482
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 TTACCATATGTTGTCTGTGGAAGTGAAGGTAATTCACAGAGTCAATCGACAACTCCAG 120
Qy 483 AAGCTGCTTAAGTCTTAAGGTTGAAGATATCTCAAGTAGAGCCGCAAGTGAAGCGGTGAG 542
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 AAGCTGCTTAAGTCTTAAGGATCTATCTAAGAGAGCTTCAAGTGAAGAGAG 180

```

```

Qy 543 CTGATGCTGTAAAGTAGTACATCCAGATCCAGGCGCAGAGATCATGTGGCTAGTT 602
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 CTCATGCTTATTAACATCAATGTTGAACCGATCAACGTGCTGAGTCAATGTTTAGCT 240
Qy 603 GACACATTC-AAGAGAGAGTGTAGATATAGCGGAACATGATGATGATAGAGGTAC 661
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 AATGTTTCAAGAGCAAAAGTTGTGATATTTCTTGAGAACAGCTTAACCTGAGAGTAAC 300
Qy 662 T-GGAGATCTCGAAAAAT 679
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 TGGAGATCTCTGGAAAAAT 319

```

RESULT 15

```

US-09-923-876-1764
; Sequence 1764, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
; APPLICANT: Laligudi, Raghunath V.
; APPLICANT: Kamigaki, Laura Y. (ItO)
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
; FILE REFERENCE: PL-0012-1 CON
; CURRENT APPLICATION NUMBER: US/09/923,876
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: 09/298,329
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/085,331
; PRIOR FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 6332
; SOFTWARE: PERL Program
; SEQ ID NO 1764
; LENGTH: 233
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20020013958A1 700159462H1
; LOCATION: 163, 166, 170-171, 173
; OTHER INFORMATION: a, t, c, g, or other
US-09-923-876-1764

```

```

Query Match      7.0%: Score 116.8; DB 10; Length 233;
Best Local Similarity 70.8%; Pred. No. 2.9e-26;
Matches 165; Conservative 0; Mismatches 67; Indels 1; Gaps 1;

```

```

Qy 1003 CTCTATCATTCCTGTGAATGATATTCACAGAGTCTTATATATGTGACTGCTTTTCG 1062
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 CTCTCTCATTCCTGTGAATGACTGCTGCTGCTCTCAACATTTGAACAGAGCTTTTG 60
Qy 1063 CTCGAAGGGGATACATATTCAGAGCTTGGCCGTGACATGCTGMAACAGAGGCATT 1122
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 CTCGAGGGGCTACATATACAGAGCTTCTGCTGTTGGCCAGCTGAGAGAGGACATT 120
Qy 1123 CAGCATTAACAAGTTATACCTGCAACAGATGAATGCGTCAAGAAATGCTGACAGAAC 1182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 CCGGATTTACAACAGTTGTTCCGTGTAATTCATTTGANGANGTTNNNCAGACAGC 180
Qy 1183 TTTACAACCTGATATGATGATGAGTCCATGATCTTACTATTTGCCATTT 1235
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 TTTACAAGCTTATGATGTGACAGA-ATTCAAGACATTAACCACTACCTCTTT 232

```

```

Search completed: March 9, 2003, 22:59:12
Job time: 160 secs

```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 09:09:40 ; Search time 17 seconds
(without alignments)
1217.974 Million cell updates/sec

Title: US-09-997-900-2

Perfect score: 2444
Sequence: 1 MAATSVSSPSIRCLRSACS.....ARESGVDSKYLGRYFLLTG 491

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 188354 seqs, 42170167 residues

Total number of hits satisfying chosen parameters: 188354

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCR_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCITUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2444	100.0	491	US-09-997-900-2	Sequence 2, Appli
2	1692	69.2	476	US-09-732-618-18	Sequence 18, Appli
3	1493.5	61.1	483	US-09-732-618-10	Sequence 10, Appli
4	1476	60.4	449	US-09-732-618-24	Sequence 24, Appli
5	1476	60.4	449	US-09-732-618-25	Sequence 25, Appli
6	1395	57.1	365	US-09-732-618-16	Sequence 16, Appli
7	1302	53.3	350	US-09-732-618-6	Sequence 6, Appli
8	522	20.4	203	US-09-732-618-2	Sequence 2, Appli
9	499	20.4	133	US-09-732-618-22	Sequence 22, Appli
10	470.5	19.3	156	US-09-732-618-4	Sequence 4, Appli
11	436	17.8	156	US-09-732-618-12	Sequence 12, Appli
12	372.5	15.2	144	US-09-732-618-14	Sequence 14, Appli
13	329.5	13.5	135	US-09-732-618-20	Sequence 20, Appli
14	307	12.6	172	US-09-738-626-6958	Sequence 6958, Ap
15	305	12.5	166	US-09-815-242-13372	Sequence 13372, A
16	305	12.5	166	US-09-815-242-13595	Sequence 13595, A
17	277.5	11.4	163	US-09-761-782-2	Sequence 2, Appli
18	277	11.3	163	US-09-815-242-12044	Sequence 12044, A
19	276.5	11.3	163	US-09-815-242-10023	Sequence 10023, A

20	271.5	11.1	173	10	US-09-815-242-14018	Sequence 14018, A
21	269.5	11.0	173	10	US-09-815-242-11742	Sequence 11742, A
22	268.5	11.0	163	10	US-09-815-242-11256	Sequence 11256, A
23	267.5	10.9	163	10	US-09-815-242-5042	Sequence 5042, Ap
24	138	5.6	828	9	US-09-732-618-8	Sequence 8, Appli
25	135	5.5	828	9	US-09-738-626-5038	Sequence 5038, Ap
26	117	4.5	495	10	US-09-815-242-10315	Sequence 10315, A
27	107	4.4	500	10	US-09-815-242-13764	Sequence 13764, A
28	106.5	4.4	5215	9	US-09-860-846-2	Sequence 2, Appli
29	106.5	4.4	5215	9	US-09-861-289-2	Sequence 2, Appli
30	105.5	4.3	503	10	US-09-815-242-11400	Sequence 11400, A
31	105.5	4.3	3503	9	US-10-108-605-237	Sequence 237, App
32	105	4.3	495	10	US-09-815-242-5040	Sequence 5040, Ap
33	105	4.3	495	10	US-09-815-242-11716	Sequence 11716, A
34	103.5	4.2	978	10	US-09-893-817-4	Sequence 4, Appli
35	103.5	4.2	978	10	US-09-893-817-8	Sequence 8, Appli
36	102.5	4.2	96	10	US-09-741-669-362	Sequence 362, App
37	102.5	4.2	737	10	US-09-815-242-13699	Sequence 13699, A
38	101	4.1	84	10	US-09-815-242-5915	Sequence 5915, Ap
39	101	4.1	84	10	US-09-815-242-12930	Sequence 12930, A
40	101	4.1	84	10	US-09-815-242-13161	Sequence 13161, A
41	101	4.1	449	10	US-09-815-242-12423	Sequence 12423, A
42	101	4.1	452	10	US-09-815-242-5585	Sequence 5585, Ap
43	101	4.1	608	9	US-09-738-626-5524	Sequence 5524, Ap
44	100.5	4.1	810	9	US-09-815-242-11397	Sequence 11397, A
45	99.5	4.1	493	10	US-09-815-242-12052	Sequence 12052, A

ALIGNMENTS

```
RESULT 1
US-09-997-900-2
Sequence 2, Application US/09997900
; Patent No. US20020053098A1
; GENERAL INFORMATION:
; APPLICANT: Kakefuda, Genchi
; APPLICANT: Costello, Colleen
; APPLICANT: Sun, Ming
; APPLICANT: Hu, Weiming
; TITLE OF INVENTION: Genes and Vectors for Confering Herbicide Resistance
; FILE REFERENCE: 043753/241148 (5849-20A)
; CURRENT APPLICATION NUMBER: US/09/997,900
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: 60/106,239
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: 09/426,568
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2:
; LENGTH: 491
; TYPE: PRT
; ORGANISM: Arabidopsis sp.
US-09-997-900-2
```

```
Query Match      100.0%; Score 2444; DB 10; Length 491:
Best Local Similarity 100.0%; Pred. No. 4.6e-203;
Matches 491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAATSVSSPSIRCLRSACSDSSPALVSTRVSFPAKISYLSGISHRDDEMKRREGFV 60
DB 1 MAATSVSSPSIRCLRSACSDSSPALVSTRVSFPAKISYLSGISHRDDEMKRREGFV 60
QY 61 RSVDGKISDASESEASATPKSKVRKHTISVFGESEGINIRIAGVFARAGYNISLAVG 120
DB 61 RSVDGKISDASESEASATPKSKVRKHTISVFGESEGINIRIAGVFARAGYNISLAVG 120
QY 121 LNRDKALFTIIVCGTFRVLOQVTELOKLVNVLKEDISSEPOVERELMLVYVNAHPER 180
DB 121 LNRDKALFTIIVCGTFRVLOQVTELOKLVNVLKEDISSEPOVERELMLVYVNAHPER 180
```



```
RESULT 4
US-09-732-618-24
; Sequence 24, Application US/09732618
; Patent No. US20010044939A1
; GENERAL INFORMATION:
; APPLICANT: Abell, Lynn
; APPLICANT: Falco, Carl
; APPLICANT: Famodu, Omolayo O.
; TITLE OF INVENTION: Small Subunit of Plant Acetolactate Synthase
; FILE REFERENCE: B01435 US NA
; CURRENT FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/174,437
; PRIOR FILING DATE: 2000-01-04
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: Microsoft Office 97
; SEQ ID NO: 24
; LENGTH: 449
; TYPE: PRT
; ORGANISM: Nicotiana glauca
US-09-732-618-24

Query Match          60.4%; Score 1476; DB 10; Length 449;
Best Local Similarity 79.1%; Pred. No. 1.6e-119;
Matches 291; Conservative 35; Mismatches 42; Indels 0; Gaps 0;

QY 99 MINRIAGFARRGYNIESLAVGLNRKALFTTVVCTERVLOQVTELOKLVNVLKVEDI 158
DB 37 MINRIAGFARRGYNIESLAVGLNRKALFTTVVCTERVLOQVTELOKLVNVLKVEDI 96
QY 159 SSEPOVERELMLKVNAPESRAEIMLVDFRFRVVDIAEHALTTEVGDGPKMAVER 218
DB 97 SSEPOVERELMLKISADPKYRAEVMVLVDYFRRAKIVDSQSLTEVGDGPKMAVOR 156
QY 219 NLKFOIREIVRTGKIALRREKMGATAPFWRPSAASYPDLKQAPVSVLRSSKGAIVPQ 278
DB 157 NLKFOIREIVRTGKIALRREKMGATAPFWRPSAASYPDLKQAPVSVLRSSKGAIVPQ 216
QY 279 KETSAGDVPYPPFDPKVRHRLDAHMGILNDEDTSGLSRSHLSLLVNDIPGVNIYVG 338
DB 217 SMSMAGDVPYPPFDPKVRHRLDAHMGILNDEDTSGLSRSHLSLLVNDIPGVNIYVG 276
QY 339 VFARRGYNIOSLAVGAEETKISRITTVIPATDESYSKLVQOLYKLVNDEHVDLTHLFP 398
DB 277 VFARRGYNIOSLAVGAEETKISRITTVIPATDESYSKLVQOLYKLVNDEHVDLTHLFP 336
QY 399 SEBELMLIKIAVNAARVDLIASIFRAKAVDSHTTTLQLTGDLKRVALLQRLLEPY 458
DB 337 SEBELMLIKIAVNAARVDLIASIFRAKAVDSHTTTLQLTGDLKRVALLQRLLEPY 396
QY 459 GICEVART 466
DB 397 GICEVART 404

RESULT 5
US-09-732-618-25
; Sequence 25, Application US/09732618
; Patent No. US20010044939A1
; GENERAL INFORMATION:
; APPLICANT: Abell, Lynn
; APPLICANT: Falco, Carl
; APPLICANT: Famodu, Omolayo O.
; TITLE OF INVENTION: Small Subunit of Plant Acetolactate Synthase
; FILE REFERENCE: B01435 US NA
; CURRENT FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/174,437
; PRIOR FILING DATE: 2000-01-04
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: Microsoft Office 97
; SEQ ID NO: 25
```

```
; LENGTH: 449
; TYPE: PRT
; ORGANISM: Nicotiana glauca
US-09-732-618-25

Query Match          60.4%; Score 1476; DB 10; Length 449;
Best Local Similarity 79.1%; Pred. No. 1.6e-119;
Matches 291; Conservative 35; Mismatches 42; Indels 0; Gaps 0;

QY 99 MINRIAGFARRGYNIESLAVGLNRKALFTTVVCTERVLOQVTELOKLVNVLKVEDI 158
DB 37 MINRIAGFARRGYNIESLAVGLNRKALFTTVVCTERVLOQVTELOKLVNVLKVEDI 96
QY 159 SSEPOVERELMLKVNAPESRAEIMLVDFRFRVVDIAEHALTTEVGDGPKMAVER 218
DB 97 SSEPOVERELMLKISADPKYRAEVMVLVDYFRRAKIVDSQSLTEVGDGPKMAVOR 156
QY 219 NLKFOIREIVRTGKIALRREKMGATAPFWRPSAASYPDLKQAPVSVLRSSKGAIVPQ 278
DB 157 NLKFOIREIVRTGKIALRREKMGATAPFWRPSAASYPDLKQAPVSVLRSSKGAIVPQ 216
QY 279 KETSAGDVPYPPFDPKVRHRLDAHMGILNDEDTSGLSRSHLSLLVNDIPGVNIYVG 338
DB 217 SMSMAGDVPYPPFDPKVRHRLDAHMGILNDEDTSGLSRSHLSLLVNDIPGVNIYVG 276
QY 339 VFARRGYNIOSLAVGAEETKISRITTVIPATDESYSKLVQOLYKLVNDEHVDLTHLFP 398
DB 277 VFARRGYNIOSLAVGAEETKISRITTVIPATDESYSKLVQOLYKLVNDEHVDLTHLFP 336
QY 399 SEBELMLIKIAVNAARVDLIASIFRAKAVDSHTTTLQLTGDLKRVALLQRLLEPY 458
DB 337 SEBELMLIKIAVNAARVDLIASIFRAKAVDSHTTTLQLTGDLKRVALLQRLLEPY 396
QY 459 GICEVART 466
DB 397 GICEVART 404

RESULT 6
US-09-732-618-16
; Sequence 16, Application US/09732618
; Patent No. US20010044939A1
; GENERAL INFORMATION:
; APPLICANT: Abell, Lynn
; APPLICANT: Falco, Carl
; APPLICANT: Famodu, Omolayo O.
; TITLE OF INVENTION: Small Subunit of Plant Acetolactate Synthase
; FILE REFERENCE: B01435 US NA
; CURRENT FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/174,437
; PRIOR FILING DATE: 2000-01-04
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: Microsoft Office 97
; SEQ ID NO: 16
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-732-618-16

Query Match          57.1%; Score 1395; DB 10; Length 365;
Best Local Similarity 75.6%; Pred. No. 1.1e-112;
Matches 276; Conservative 38; Mismatches 49; Indels 2; Gaps 1;

QY 127 LFTIVVCTERVLOQVTELOKLVNVLKVEDISSEPOVERELMLKVNAPESRAEIMML 186
DB 1 MFTIVVCTERVLOQVTELOKLVNVLKVEDISSEPOVERELMLKINVEPDQRPVAVL 60
QY 187 VDFRFRVVDIAEHALTTEVGDGPKMAVERLKFQFIREIVRTGKIALRREKMGATAP 246
DB 61 VDFRFRVVDIAEHALTTEVGDGPKMAVERLKFQFIREIVRTGKIALRREKMGATAP 120
QY 247 FWRPSAASYPDLKQAPVSVLRSSKGAIVP--QKETSAGDVPYPPFDPKVRHRLDA 304
```



```

QY 168 IMLVYNHPESSRAIMLVDTFERRAVDIAHALTIEVYGPCKIAVERNLKFKQIRE 227
Db 63 IMLILNVEPQORADVMYFVANYFRAKVYDISENSLTEVYGPCKIAAQRNLKFKGIEE 122
QY 228 IVRTGKIALRR 238
Db 123 ICRTGKIALRRQ 133

RESULT 10
US-09-732-618-4
; Sequence 4, Application US/09732618
; Patent No. US20010044939A1
; GENERAL INFORMATION:
; APPLICANT: Abell, Lynn
; APPLICANT: Falco, Carl
; APPLICANT: Farnodu, Omolayo O.
; TITLE OF INVENTION: Small Subunit of Plant Acetolactate Synthase
; FILE REFERENCE: B01435 US NA
; CURRENT APPLICATION NUMBER: US/09/732,618
; CURRENT FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/114,437
; PRIOR FILING DATE: 2000-01-04
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 4
; LENGTH: 156
; TYPE: PRN
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (35)
; NAME/KEY: UNSURE
; LOCATION: (43)
; NAME/KEY: UNSURE
; LOCATION: (55)
; NAME/KEY: UNSURE
; LOCATION: (57)..(58)
; NAME/KEY: UNSURE
; LOCATION: (64)
; NAME/KEY: UNSURE
; LOCATION: (66)
; NAME/KEY: UNSURE
; LOCATION: (119)
; NAME/KEY: UNSURE
; LOCATION: (130)
; NAME/KEY: UNSURE
; LOCATION: (149)
; NAME/KEY: UNSURE
; LOCATION: (151)
;
US-09-732-618-4

Query Match 19.3%; Score 470.5; DB 10; Length 156;
Best Local Similarity 63.5%; Pred. No. 2.2e-33;
Matches 99; Conservative 15; Mismatches 37; Indels 5; Gaps 2

QY 141 OYIELOKLVNVLKVEDISSPEQVREMLLVKNHNPESRAITMLVDTFFRARVVDIAEH 200
Db 1 OYIELOKLVNHNHVEDLSKEPVERELMLIKLNHPDQRPXMYLVDTFFRAKVXDXXER 60
QY 201 ALTIEVTDGPKMIAVERNLKFKQIREIVRTGKIALRREKMGATAPFRFSSAASYPDLKE 260
Db 61 TLTXKXAGDPCKIAVORNLKFKGIEICRTGKIALRREKIGATARFWRFSASYPDLME 120
QY 261 QAPVSVLSSKR--GAIVPOKETYSAGGDVYVEPP 293
Db 121 ALPKKPLTSXNKTVNGSFV--RPSNAGGDVXPXESY 154

RESULT 11
US-09-732-618-12
; Sequence 12, Application US/09732618

```

[illegible]

```

RESULT 12
US-09-732-618-14
; Sequence 14, Application US/09732618
; Patent No. US20010044939A1
; GENERAL INFORMATION:
; APPLICANT: Abell, Lynn
; APPLICANT: Falco, Carl
; APPLICANT: Famodu, Omolayo O.
; TITLE OF INVENTION: Small Subunit of Plant Acetolactate Synthase
; FILE REFERENCE: B81435 US NA
; CURRENT APPLICATION NUMBER: US/09/732,618
; PRIORITY FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/174,437
; PRIOR FILING DATE: 2000-01-04
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 14
; LENGTH: 144
; TYPE: PRY
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (114)
; NAME/KEY: UNSURE
; LOCATION: (137)..(138)
; NAME/KEY: UNSURE
; LOCATION: (142)
; NAME/KEY: UNSURE
; LOCATION: (144)
; US-09-732-618-14
Query Match      15.2%; Score 372.5; DB 10; Length 144;
Best Local Similarity   60.6%; Pred.No. 5.6e-25;
Matches    86; Conservative 15; Mismatches 24; Indels 17; Gaps 5;
Oryza sativa L. cv. IR64 [GenBank: U00180]

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 9, 2003, 18:39:32 ; Search time 4658 Seconds

(without alignments)
10452.768 Million cell updates/sec

Title: US-09-997-900-1

Perfect score: 1673
Sequence: 1 gctctctcagtagcaaaaa.....tgatttggcttcattcg 1673

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

(GenEmbl):*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pac:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_on:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1673	100.0	1673	6	AR194786	AR194786 Sequence
2	1644.6	98.3	1675	8	AY058881	AY058881 Arabidops
3	1474.4	88.1	1476	8	AY103307	AY103307 Arabidops
4	665.8	39.8	1860	8	NP1234901	NP1234901 Nicotiana
5	420	25.1	100665	8	AC006533	AC006533 Arabidops
6	419	25.0	4895	6	AF233069	AF233069 Arabidops
7	156.6	9.4	6982	8	AF233069	AF233069 Arabidops
8	153.8	9.2	164921	8	AF022186	AF022186 Cyanidium
9	150.8	9.0	300450	1	AP005371	AP005371 Thermosyr
10	149.8	9.0	11318	1	AE013291	AE013291 Methanosa
11	147.8	8.8	191028	8	PP38804	PP38804 Porphyra pu
12	147.2	8.8	7550	8	D63676	D63676 Cyanidium c
13	146	8.7	11759	1	AE011091	AE011091 Methanosa
14	145	8.7	121524	8	AF041468	AF041468 Guillardia
15	145	8.7	342850	1	AP003597	AP003597 Nostoc sp
16	137.4	8.2	8200	8	D63675	D63675 Cyanidiosch
17	132.2	7.9	105795	1	SYCCPNC	SYCCPNC
18	130.6	7.8	116302	2	AC129715	AC129715 Medicago
19	128	7.7	82001	8	AB005242	AB005242 Arabidops
20	124.8	7.5	10481	1	AE012976	AE012976 Thermosyr
21	123.8	7.4	104383	2	AP004227	AP004227 Oryza sat
22	123.8	7.4	138556	2	AP004881	AP004881 Oryza sat
23	118.4	7.1	13104	1	AE000985	AE000985 Archaeog
24	111.4	6.7	12156	1	AE000759	AE000759 Aquifex a
25	104.8	6.3	15939	1	AE007128	AE007128 Mycobacte
26	104.8	6.3	70287	1	MTY012	MTY012 Mycobacte
27	104.4	6.2	10408	1	AE007812	AE007812 Clostridi
28	103.4	6.2	13072	1	U67473	U67473 Methanococc
29	98.4	5.9	44882	1	MLCB637	MLCB637 Mycobacteri
30	98.4	5.8	348950	1	AF118061	AF118061 Mycobacte
31	97	5.8	2265	1	AE000905	AE000905 Methanoc
32	92	5.5	11643	1	AX100789	AX100789 Methanoba
33	91	5.4	525	6	AX100789	AX100789 Sequence
34	91	5.4	3291	6	AX100814	AX100814 Sequence
35	91	5.4	4281	1	BACILNB	BACILNB
36	91	5.4	10426	6	AX100838	AX100838 Streptoco
37	91	5.4	89009	1	BS275208	BS275208 B. subtilis
38	91	5.4	218410	1	BSUB0015	BSUB0015 Bacillus su
39	90.4	5.4	6211	6	BD003688	BD003688 Polynucle
40	90.4	5.4	10762	1	AE007355	AE007355 Streptoco
41	90.4	5.4	11274	1	AE008420	AE008420 Streptoco
42	90.4	5.4	11135	2	SPNEU1906	SPNEU1906 Streptoco
43	89	5.3	4210	1	MSGILVB	MSGILVB
44	87.4	5.2	3527	1	MAU35458	MAU35458 Methanococ
45	85.4	5.1	10029	1	AE012835	AE012835 Chlorobiu

ALIGNMENTS

RESULT 1
AR194786
LOCUS AR194786 1673 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6348643.
ACCESSION AR194786
VERSION AR194786.1 GI:20241378
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1673)
TITLE Kakefuda, G., Costello, C., Sun, M. and Hu, W.
DNA sequences encoding the arabinidops acetylhydroxy-acid synthase
small subunit and methods of use
JOURNAL Patent: US 6348643-A 1 19-FEB-2002;

FEATURES		Location/Qualifiers			
Source	1. .1673	/organism="unknown"			
BASE COUNT	466 a	328 c	418 g	461 t	
ORIGIN					
Query Match	100.0%	Score 1673;	DB 6;	Length 1673;	
Best Local Similarity	100.0%	Pred. No. 0;			
Matches 1673;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
Oy	1	GNCTCTTCAGTAGAGAAAAAACCTTCGGCTTCGTCGTCACATGGGGGCATTTCGTAA	60		
Db	1	GCTCTCTTCAGTAGAGAAAAAACCTTCGGCTTCGTCGTCACATGGGGGCATTTCGTAA	60		
Oy	61	GTTCTTCACCATCTATTCGCTGCTTGAGATCGCATGTGTCGATTCCTCCTGCTTG	120		
Db	61	GTTCTTCACCATCTATTCGCTGCTTGAGATCGCATGTGTCGATTCCTCCTGCTTG	120		
Oy	121	TATCCTCGACCGGTGATATCTCCGGCGAGATTTCAATATCTCCGTAATCTTGGC	180		
Db	121	TATCCTCGACCGGTGATATCTCCGGCGAGATTTCAATATCTCCGGTATATCTTCGC	180		
Oy	181	ACCGGGCGATAAATGGGTAGAGAAATGGAGAAATTCGTTGAAGCGTCGATGGGAAA	240		
Db	181	ACCGGGCGATAAATGGGTAGAGAAATGGAGAAATTCGTTGAAGCGTCGATGGGAAA	240		
Oy	241	TCTCTGATGCGCTCTTCTCCGAAGCTTCATCTGCGACTCCAAATCGAAGTAGAGAAC	300		
Db	241	TCTCTGATGCGCTCTTCTCCGAAGCTTCATCTGCGACTCCAAATCGAAGTAGAGAAC	300		
Oy	301	ACACATTTTCAGTATTTTGTGAGACGAAAGCGGAATGATTATAGATTCAGAGACTG	360		
Db	301	ACACATTTTCAGTATTTTGTGAGACGAAAGCGGAATGATTATAGATTCAGAGAGTGT	360		
Oy	361	TTCGAGGAGAGATACATATTGAGAGCTTCGCTGGTGGATGGAACAGACAAAGGCTC	420		
Db	361	TTCGAGGAGAGATACATATTGAGAGCTTCGCTGGTGGTGAACAGACAAAGGCTC	420		
Oy	421	TATTCACCATAGTTTCTGTGGAAGGCTACTTCAGCAGGTCATCGAGCAACTCC	480		
Db	421	TATTCACCATAGTTTCTGTGGAAGGCTACTTCAGCAGGTCATCGAGCAACTCC	480		
Oy	481	AGAACTCGTTATGTCTTCTTAAAGGTTGAAGATATCTCAAGTACGCCGCAAGTGAGCGTG	540		
Db	481	AGAACTCGTTATGTCTTCTTAAAGGTTGAAGATATCTCAAGTACGCCGCAAGTGAGCGTG	540		
Oy	541	AGCTATGCTTTAAAGGAATGACATCCGAATCCAGGGCAGAGATCATGTGGCTAG	600		
Db	541	AGCTATGCTTTAAAGGAATGACATCCGAATCCAGGGCAGAGATCATGTGGCTAG	600		
Oy	601	TTGACACATTCAGACCAAGAGTTGTAGATATAGCGGAACATCATTTGACTATCGAGTAA	660		
Db	601	TTGACACATTCAGACCAAGAGTTGTAGATATAGCGGAACATCATTTGACTATCGAGTAA	660		
Oy	661	CTGGAGATCTTGGAAAAATGATTCGCTGTAGAAAATTTGAAAAAGTTTCAGATCAGAG	720		
Db	661	CTGGAGATCTTGGAAAAATGATTCGCTGTAGAAAATTTGAAAAAGTTTCAGATCAGAG	720		
Oy	721	AGATTGTAAAGCAGAGAAAGATAGACAGTGAAGGAAAGATGGGTGCACTGCTCCAT	780		
Db	721	AGATTGTAAAGCAGAGAAAGATAGACAGTGAAGGAAAGATGGGTGCACTGCTCCAT	780		
Oy	781	TTTGCGCATTTTTCACAGCATCTATCCGATCTCAAGGAGACAAGCGCTGTAGTGTTC	840		
Db	781	TTTGCGCATTTTTCACAGCATCTATCCGATCTCAAGGAGACAAGCGCTGTAGTGTTC	840		
Oy	841	TTTGAAGTAGCAAAAAAGAGCCATTGTCCTCAAAAAGGAACATCAGAGGGGAGATG	900		
Db	841	TTTGAAGTAGCAAAAAAGAGCCATTGTCCTCAAAAAGGAACATCAGAGGGGAGATG	900		
Oy	901	TTTATCCCGGTGAGCATTTTGTGACCCCAAGGTACATGTAATTCGACGCTCACTGGG	960		
Db	901	TTTATCCCGGTGAGCATTTTGTGACCCCAAGGTACATGTAATTCGACGCTCACTGGG	960		

[illegible]

REFERENCE 2 (bases 1 to 1675)
AUTHORS Shin,P., Chen,H., Cheuk,R., Kim,C.J., Koesema,E., Meyers,M.C.,
Banh,J., Bowser,L., Carninci,P., Dale,J.M., Goldsmith,A.D.,
Hayashizaki,Y., Ishida,J., Jiang,P.X., Jones,T., Kamiya,A.,
Karlin-Neumann,G., Kawai,J., Lam,B., Lee,J.M., Lin,J., Liu,S.X.,
Miranda,M., Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J.,
Pham,P.K., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A.,
Tang,C.C., Toriumi,M., Yamada,K., Yamamura,Y., Yu,G., Yu,S.,
Shiozaki,K., Davis,R.W., Theologis,A. and Ecker,J.R.
TITLE Direct Submission
JOURNAL Submitted (05-SEP-2001) Salk Institute Genomic Analysis Laboratory
(SIGNAL), Plant Biology Laboratory, The Salk Institute for
Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
USA

FEATURES
source Location/Qualifiers
1. 1675
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="2"
/clone="RAF109-09-C24(R14179)"
/note="ecotype: Columbia"
50..1525
/note="putative acetolactate synthase"
/codon_start=1
/product="At2g31810/F20M17.15"
/protein_id="AAL24267.1"
/db_xref="GI:16604523"
/translation="MAAISVSSSPSIRCLRSACSDSPALVSSTVSPAPKISYLSGI
SSHRDEMRKHEGVRSYDKISDPSSEASATPKSVKRHTTSVFEDSGMINK
IAGVFARKGYNIESLAVGLNRKALFTLVCGTEVLDQVIEQLKLVNLVVEDIS
EPOVREMLVNVNHPESRAIMLVDFEFARVVDIEHALTIEVTPGPKMIAVER
NLKFOIREIVTGTALREKEMATAPFRSAASYDLKQAPVSVLRSSKGAIV
POKTSAGDVYVPEPFDPKYHRLIDAHMLGLTDESKSLSLVNDIPGLN
ITGVFARGVYIOSLAVGHATKGISRTTTPATPDESVSKLNOILKLVNHHVHD
LTHLPFSREMLIKIYNVNAARRDVLDIASFRAKADVSHITTLQLTGLDLMNVA
LORLEPFGICVATGRVALRESGVDSKILRGVSFPLTG"

CDS
Best Local Similarity 98.3%; Score 1644.6; DB 8; Length 1675;
Pred. No. 0;
Matches 1660; Conservative 0; Mismatches 4; Indels 3; Gaps 1;

BASE COUNT 469 a 331 c 417 g 458 t
ORIGIN

Db 429 TATTACCAATAGTTGCTGTGGAAGTGAAGGGGTACTTCAGCAGGTCAATCGAACCACTCC 488
Oy 481 AGAAGCTCGTAATATGTTCTAAAGGTTGAAGATATCCAGAGTGAGCGCGCAAGTGAAGCTG 540
Db 489 AGAAGCTCGTAATATGTTCTAAAGGTTGAAGATATCCAGAGTGAGCGCGCAAGTGAAGCTG 548
Oy 541 AGCTGATGCTTGTAAAGTGAATGCACATCCAGAAATCCAGGCGAGATCATGTGGCTAG 600
Db 549 AGCTGATGCTTGTAAAGTGAATGCACATCCAGAAATCCAGGCGAGATCATGTGGCTAG 608
Oy 601 TTGACACATTCAGACAGAGAGTTGTAGATATAGCGGAACATGCTATTGACTATCGAGCTAA 660
Db 609 TTGACACATTCAGACAGAGAGTTGTAGATATAGCGGAACATGCTATTGACTATCGAGCTAA 668
Oy 661 CTGAGATCCCTGGAAGAAATGATTCGTGTGAAGAAATTTGAAAGTTTCAGATCAGAG 720
Db 669 CTGAGATCCCTGGAAGAAATGATTCGTGTGAAGAAATTTGAAAGTTTCAGATCAGAG 728
Oy 721 AGATTGTAGAGACAGGAAGATAGCACTAGAGAGGAGAAAGATGGGTGCACACTGCTCAT 780
Db 729 AGATTGTAGAGACAGGAAGATAGCACTAGAGAGGAGAAAGATGGGTGCACACTGCTCAT 788
Oy 781 TTGGCGGATTTTCAGCAGCATCTCTATCCAGATCTCAAGAGCAGCGCTGTAGTGTTC 840
Db 789 TTGGCGGATTTTCAGCAGCATCTCTATCCAGATCTCAAGAGCAGCGCTGTAGTGTTC 848
Oy 841 TTGCAAGTGAACAAAAAGAGAGCATCTGTCCTCAAAAGAACATCAGAGGGGAGATG 900
Db 849 TTGCAAGTGAACAAAAAGAGAGCATCTGTCCTCAAAAGAACATCAGAGGGGAGATG 908
Oy 901 TTATATCCCGTTGAGCCATTTTGTGACCCCAAGGTACATGTTATTCGAGCGCTACACTGG 960
Db 909 TTATATCCCGTTGAGCCATTTTGTGACCCCAAGGTACATGTTATTCGAGCGCTACACTGG 968
Oy 961 GACTTTCTACTGACGAGATAGCATGAGTACGCTGCTGATCTATGATCTTGTAA 1020
Db 969 GACTTTCTACTGACGAGATAGCATGAGTACGCTGCTGATCTATGATCTTGTAA 1028
Oy 1021 ATGATATTCAGAGATGTTCTTAATATGTGATGATCTGTTTTCGCTGCAAGGGATTAACAATA 1080
Db 1029 ATGATATTCAGAGATGTTCTTAATATGTGATGATCTGTTTTCGCTGCAAGGGATTAACAATA 1088
Oy 1081 TCCAGAGCTTTGGCGGTAGGACATGCTGAAACCAGGGCATTTTACGCATTACAAAGTTA 1140
Db 1089 TCCAGAGCTTTGGCGGTAGGACATGCTGAAACCAGGGCATTTTACGCATTACAAAGTTA 1148
Oy 1141 TACCTGCAACAGATGATGCGTCAAGCAAAATGCTGACAACTTTACAACTGTAGATG 1200
Db 1149 TACCTGCAACAGATGATGCGTCAAGCAAAATGCTGACAACTTTACAACTGTAGATG 1208
Oy 1201 TGCATAGAGTCCATGATCTTACATCTTTTGGCATTTTCTGAAAGAGACATGATCTGATTA 1260
Db 1209 TGCATAGAGTCCATGATCTTACATCTTTTGGCATTTTCTGAAAGAGACATGATCTGATTA 1268
Oy 1261 AGATTCCCGTGAACGCTGCTGCTAGAGAGATGCTCTGACATTTGCTAGTATTTTCAGGG 1320
Db 1269 AGATTCCCGTGAACGCTGCTGCTAGAGAGATGCTCTGACATTTGCTAGTATTTTCAGGG 1328
Oy 1321 CTAAAGCTGTTGACGATCTGATCACACAATTAATCTTTGACGCTTACTGGGGATCTAGACA 1380
Db 1329 CTAAAGCTGTTGACGATCTGATCACACAATTAATCTTTGACGCTTACTGGGGATCTAGACA 1388
Oy 1381 AGATTGCTGACGCAAGAGTTATTTGAGGCCCTATGCTATATGATGAGGTTGCAAGAACCG 1440
Db 1389 AGATTGCTGACGCAAGAGTTATTTGAGGCCCTATGCTATATGATGAGGTTGCAAGAACCG 1448
Oy 1441 GTGCTGTGCAATGGCTCGTAATCGGGAGTGAATCCAAAGTCTTCTGCTGATTAATCTCT 1500
Db 1449 GTGCTGTGCAATGGCTCGTAATCGGGAGTGAATCCAAAGTCTTCTGCTGATTAATCTCT 1508
Oy 1501 TTCTTTTAAACAGGCTTAACCGCTTGCAGAGTGCATTCATGCAACATGAGAACTTTTGAAG 1560
Db 1509 TTCTTTTAAACAGGCTTAACCGCTTGCAGAGTGCATTCATGCAACATGAGAACTTTTGAAG 1568

[illegible]

FEATURES
source

The annotation of this entry was produced with considerable contributions from Stephane Kombauts and Pierre Rouze, Department of Genetics, University of Ghent, Leuvenagkstrat 35, 9000 Ghent BE. E-mail: strom@gengemp.rug.ac.be, piroze@gengemp.rug.ac.be A more detailed annotation of this entry and other sequences from the long arm of chromosome 4 can be viewed at: <http://websrv.mips.biochem.mpg.de/proj/thal/> this fragment has an overlap with ATAA22 at the 3' end.

misc_feature 1..222
/note="overlap with BAC clone F22D22 (AC006223:102973..103194)."

gene 1..127
/gene="At2g31955"
/note="F20M17..27"
<1..>127
/gene="At2g31955"

CDs 1..127
/gene="At2g31955"
/note="Protein sequence is in conflict with the conceptual translation"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA015275.1"
/db_xref="GI:20197843"
/translation="NSNFLVNPPTLLILTSLQWWSISPTQVCFEELISCSy"
231..2290
/gene="At2g31950"
/note="F20M17..1; supported by full length cDNA: Ceres:33192"
join(231..824,922..1089,1185..1376,1551..1619,1720..1815,1985..2290)
/gene="At2g31950"
join(357..824,922..1089,1185..1376,1551..1619,1720..1803)
/gene="At2g31950"
/codon_start=1
/product="molybdopter in synthase (CMX2)"
/protein_id="AA015276.1"
/db_xref="GI:20197844"
/translation="MLIDKFGRLHTYLRISLTERCNLRQCYCMPSGEVELTPKPOLLS
QSEIYRLAGLSVAGVKNIRLTGEPYVRKDIEICLOSLSLGKLNLAITNGITLA
KRLPLKRCGSDLSNISLDTLPAPKKEFLRRRGHDVRKMSIDTAEIGNPKPVKCA
IMRGINDDEICDPEVLTNRDKRINVRPEMPDPGSAWNNKKLVPVAFVMDKVKRPS
IKRMODHTEYAKNFTTIDHGCGSVSFTSMTNERFCAGCNRLRLADGNKVCYLCFSE
VSLRDLPTSGADDEALRELIICAAVKKRKAHAGMLDIAKTANRPMHIGS"

repeat_region complement(3384..3421)
/rpt_family="AT-rich"
complement(4328..4852)
/gene="At2g31945"
/note="F20M17.2; supported by full length cDNA: Ceres:258917"
complement(<4328..>4852)
/gene="At2g31945"
complement(4460..4747)
/gene="At2g31945"

CDs /codon_start=1
/product="Expressed protein"
/protein_id="AA015278.1"
/db_xref="GI:20197846"
/translation="MDGGIGMLMKVALPALVOGLIYVLLIKSSRVAFARNSIKRAYS
FRPMKSVISRIITLALDIPACGDMSPSSSSGSSSDDEAATTTTSSPS"
complement(5476..5638)
/gene="At2g31940"
/note="F20M17..26; predicted by genscan"

gene

CDs

gene

CDs

gene

mRNA
CDS
repeat_region
repeat_region
gene
mRNA
CDS
repeat_region
repeat_region
gene
mRNA
CDS
repeat_region
repeat_region
gene
mRNA
CDS

complement(<5476..>.5838)
/gene="At2g31940"
complement(5476..5838)
/gene="At2g31940"
/codon_start=1
/product="hypothetical protein"
/protein_id="ABD32278.2"
/db_xref="GI:20197841"
translation="MGROQSGSPRYPLHLCVFVLLLMFVITSWASYEPYIEGFTQ
FLTAMASPLLILLAVLHFLESDDGVGGMWTSLIHLNERESLYRAGGRPWGAFMLVEL
FMVYSOFOGRMPELR"

complement(6202..6233)
/rpt_family="AT-rich"
complement(6743..6763)
/rpt_family="AR-rich"
complement(6858..8872)
/gene="At2g31930"
/note="P20M17.3; predicted by genescan"
complement(join(<6858..6919,8530..>.8872))
/gene="At2g31930"
complement(join(6858..6919,8530..8872))
/gene="At2g31930"
/codon_start=1
/product="hypothetical protein"
/protein_id="ABD32279.1"
/db_xref="GI:4887743"
translation="MGSEVVNPKRPFDLSKSRTRKPWSLLINDMHNAICTOSSOP
RKDRDOEDIEDIKRLNRCDEKLGEADVNNKSLLDHGFDEEGEIKPTMOLV
YKKKGGLTKFKELSQHIFRLLEFLSTRDY"

10071..12385
/gene="At2g31920"
/note="P20M17.4; predicted by genescan"
join(<10071..10640,11040..11417,11486..11577,
11668..>.12385)
/gene="At2g31920"
join(10071..10640,11040..11417,11486..11577,11668..
12385)
/gene="At2g31920"
/codon_start=1
/product="hypothetical protein"
/protein_id="ABD32280.1"
/db_xref="GI:4887744"
translation="MASLPGVLSMLINIAAGNSPSSPELLSHRSPLLOYIETVPC
LSDDDKSEAFEVKVYSDSLAAHYAAYSTDADLLRSDEIQLGFTYICGGALHKRGCP
VPVINGLHPVRKRMCVGNPSDLSDLDLLEPFGVSYPPTTKKNILNETRSLSDA
RSCGDONPPVRRRDALLSPLRSKVLIVNLISLKLTISPKEFCGSPALARK
NVKRGSPYASMASSPKDGIRKSPISKHCEFPALRSRYVPAAPISWTSPKGIS
LSKATTPVALFKPVSSHMTMSDOMSMBSGIPKTIQLGLKVSSHROYANSAVSLE
BASAENSEVFSIQAFLEJCDSTKRLSDQVAARELDYHNNTCKAHRIILTNRNN
EKSLRVVNKNATNAYSQAAYVNGEQNLNFKEPRGKGDAHDHYITVMONSXKPNKET
TSSRPSTYGKAPEPATKHCYSVDRLSKLSDSTSDELIVOVSSOMFLAKLENS
LNKGSFLVYTEETNGKESLILVHLKALNHMDLNLNRETNEKVEDLRKKLQRFLLEH
IESVIGDNYNMKRIINSI"

complement(10655..10679)
/rpt_family="(A)n"
complement(12419..14721)
/gene="At2g31910"
/note="P20M17.5"
complement(join(<12419..14247,14343..>.14721))
/gene="At2g31910"
complement(join(12419..14247,14343..14721))
/gene="At2g31910"
/codon_start=1
/product="putative Na+/H+ antiporter"
/protein_id="ABD32281.1"
/db_xref="GI:4887745"
translation="WLTFANLALVNVFLIGLIDIMRTIKDKIPIYAIVGLLA
ALLAAGADLPENGEADRILAGCWYSIAEGCTNFDPILARIADLKLIRDMQHTM
CAAYTDLTWLTEFGMAIFSKSGVRNEMRPYLASTAIPLADLYCYITOGVAMFN
NVEGQGDTHLVITLAGVIITCSLITEVCGHSTGTGFLGSIPIHHIIIRKMIEK
LHDFDSGMIMFWYITGLRADIDGYMNATVSGMAAYVTASVWKVTLSTPGCIFLR
IPLRGCLAIGALMNTKGTMALVYILNAGDTALADVITKTHTLLTLPVASKVQDLPL
ATKPKLILETVKNRTIQKHGESELCLVTCVAILPNVSGITNLIQLSNPTKSPSLNP
ATHVLEKGTSTAALLINNDEKPKRANADREVRESDOIASEPFALAEVNDNGVVOQT

Query Match	Best Local Similarity	86.8%	Pred. No. 7.7e-109;	Mismatches 0;	Indels 77;	Gaps 1;
Query 1	GTCTCTTCAGTACGCAAAAACCTTCGCTTCGTCCTGCATCTGGCGGCATTTCTGTAA	60				
Db 62208	GTCTCTTCAGTACGCAAAAACCTTCGCTTCGTCCTGCATCTGGCGGCATTTCTGTAA	62149				
Query 61	GTCTCTTCACATCTATTCGCTTCGTCAGATCGCATTCGCAATTCCTTCGCTTCG	120				
Db 62148	GTCTCTTCACATCTATTCGCTTCGTCAGATCGCATTCGCAATTCCTTCGCTTCG	62089				
Query 121	TATCCTCGACGCGGTATTCGTCCTCCGCGAGATTCATATCTCCCGTATATCTTCG	180				
Db 62088	TATCCTCGACGCGGTATTCGTCCTCCGCGAGATTCATATCTCCCGTATATCTTCG	62029				
Query 181	ACCGTGGCATTAATGGGTAAAGATTAAGAGATTCGTTAGAAAGCTCGATGGAGAA	240				
Db 62028	ACCGTGGCATTAATGGGTAAAGATTAAGAGATTCGTTAGAAAGCTCGATGGAGAA	61969				
Query 241	TCTCTGATGCTCTTCCTCCGAGAGCTTCATCGCGCATCCCAATTCGAA-----	289				
Db 61968	TCTCTGATGCTCTTCCTCCGAGAGCTTCATCGCGCATCCCAATTCGAAAGTCTGTAA	61909				
Query 290	-----	289				
Db 61908	AATATTGCTTAAGAGCTTTCTTTGCGCTTTGCTTGAATGATTTCTTGTGCAATTA	61849				
Query 290	-----GGTGGAGGAACACACAAATTTCAATTTGTTGGAGACGAAGCGGATGTTAA	343				
Db 61848	AATCAGGAGTGAAGACACACAAATTTCAATTTGTTGGAGACGAAGCGGATGTTAA	61789				
Query 344	TAGCATTCGAGAGTCTTTCGAAGAGAGATTAATTTGAGAGCTCTGCTGTAGCT	403				
Db 61788	TAGCATTCGAGAGTCTTTCGAAGAGAGATTAATTTGAGAGCTCTGCTGTAGCT	61729				
Query 404	GACAGAGACAAGGCTTATTCACCAATAGTTCTCTGTGAACTGAAAGGTAATTCAGCA	463				
Db 61728	GACAGAGACAAGGCTTATTCACCAATAGTTCTCTGTGAACTGAAAGGTAATTCAGCA	61669				
Query 464	GGTCATTCGAGCAATTCGAGAGCTGTTATGTTCTTAAAGGTTG	507				
Db 61668	GGTCATTCGAGCAATTCGAGAGCTGTTATGTTCTTAAAGGTTG	61625				
Result 6	AR194787	4895 bp	DNA	linear	PAT	20-APR-2002
LOCUS	AR194787	4895 bp	DNA	linear	PAT	20-APR-2002
DEFINITION	Sequence 3 from patent US 6348643.					
ACCESSION	AR194787					

VERSION	AR194787.1	GI:20241379
KEYWORDS	Unknown.	
SOURCE	Unknown.	
ORGANISM	Unclassified.	
REFERENCE	1 (bases 1 to 4895)	
AUTHORS	Kakefuda, G., Costello, C., Sun, M. and Hu, W.	
TITLE	DNA sequences encoding the arbidopsis acetylhydroxy-acid synthase	
JOURNAL	small subunit and methods of use	
FEATURES	Patent: US 6348643-A 3 19-FEB-2002;	
source	Location/Qualifiers	
	1..4895	
BASE COUNT	/organism="unknown"	
ORIGIN	1236 a 884 c 1011 g 1703 t 1 others	
Query Match	25.0%; Score 419; DB 6; Length 4895;	
Best Local Similarity	86.7%; Pred. No. 1e-108;	
Matches	507; Conservative 0; Mismatches 0; Indels 78; Gaps 1;	
QY	1	GTCCTTCCTCAGTAGCAAAAACCTTCGGCTTCGTCGTCGAATGCGCGGCATTCTGTAA 60
Db	717	GTCCTTCCTCAGTAGCAAAAACCTTCGGCTTCGTCGTCGTCGAATGCGCGGCATTCTGTAA 776
OY	61	GTTCTTCACCATCTATTTGGCTGCTTGAGATCGGCATGTTCCGATTCCTCTGCTCTTG 120
Db	777	GTTCTTCACCATCTATTTGGCTGCTTGAGATCGGCATGTTCCGATTCCTCTGCTCTTG 836
OY	121	TATCCTCGACGGGTATTCGTCGCCGGGAGATTCATATCTCCGGTATATCTTGGC 180
Db	837	TATCCTCGACGGGTATTCGTCGCCGGGAGATTCATATCTCCGGTATATCTTGGC 896
OY	181	ACCGTGGCGATGAATGGGTAGAGAAATGAAGATTCGTTGAGACGCGATGGGAAGA 240
Db	897	ACCGTGGCGATGAATGGGTAGAGAAATGAAGATTCGTTGAGACGCGATGGGAAGA 956
OY	241	TCTCTGATGCGCTTCTTCTCCGAGAGCTTCATCTGCACTCCAAATCGAA----- 289
Db	957	TCTCTGATGCGCTTCTTCTCCGAGAGCTTCATCTGCACTCCAAATCGAGCTGTGA 1016
OY	290	----- 289
Db	1017	TAAATTTGGCTTAAAGTCGTTCTTTGGCCCTTTCGTTGATGATTCCTTTGTGCATTA 1076
OY	290	-----GATGAGAGACACACATTTTCAGTATTTGTTGGAGACGAAAGCGGATGATTA 342
Db	1077	AAATGAGGGGTGAGAGACACACATTTTCAGTATTTGTTGGAGACGAAAGCGGATGATTA 1136
OY	343	ATAGGATTCGACAGAGTGTTCGAGAGAGATACAAATATTGAGAGTCTGCTGTGGTC 402
Db	1137	ATAGGATTCGACAGAGTGTTCGAGAGAGATACAAATATTGAGAGTCTGCTGTGGTC 1196
OY	403	TGAACAGAGACAAAGCCTCTATTCCACCATAGTGTCTGTGGAACCTAAAGGATCTTACG 462
Db	1197	TGAACAGAGACAAAGCCTCTATTCCACCATAGTGTCTGTGGAACCTAAAGGATCTTACG 1256
OY	463	AGGTCATCGAGCAACTCCAGAGCTGTTAATGTTCTTAAAGGTTG 507
Db	1257	AGGTCATCGAGCAACTCCAGAGCTGTTAATGTTCTTAAAGGTTG 1301
RESULT 7		
AF233069/c	AF233069	
LOCUS	6982 bp DNA linear	
DEFINITION	Galdieria sulphuraria maturase (matK) gene, partial cds; 508	
	ribosomal protein, ribulose-1,5-bisphosphate carboxylase/oxygenase	
	large subunit (rbcL), ribulose-1,5-bisphosphate	
	carboxylase/oxygenase small subunit (rbcS), acetylhydroxy-acid	
	synthase small subunit (ahas), and translation initiation factor	
	IF-3 (if3) genes, complete cds; and unknown genes; chloroplast	
	genes for chloroplast products.	
ACCESSION	AF233069	
VERSION	AF233069.1 GI:8925945	

KEYWORDS	FEATURES
SOURCE	Location/Qualifiers
ORGANISM	1. 6982
ORGANISM	/organism="Galdieria sulphuraria"
ORGANISM	/organelle="plastid:chloroplast"
ORGANISM	/strain="UTEX 2393"
ORGANISM	/db_xref="taxon:130081"
ORGANISM	complement(<1. 407)
ORGANISM	/gene="mark"
ORGANISM	complement(<1. 407)
ORGANISM	/gene="mark"
ORGANISM	/codon_start=1
ORGANISM	/product="maturase"
ORGANISM	/protein_id="AAE81678.1"
ORGANISM	/db_xref="GI:8925946"
ORGANISM	/translation="MKHSDIFPKKEKLEFKYLLVNIHPSLGARSVPIENMSYKNIA RADLNLVADKKYSVDYALKLEDEYINKILNRNONIAEKYRAVDLETYDKLITAKEK RFAVDNFRRYKYLFLDRLRKGHITKFGGFD."
ORGANISM	complement(552. . 860)
ORGANISM	/note="ORF339: hypothetical 12.5 kda protein"
ORGANISM	/codon_start=1
ORGANISM	/product="unknown"
ORGANISM	/protein_id="AAE81679.1"
ORGANISM	/db_xref="GI:8925947"
ORGANISM	/translation="MNTYNTNITKSDLHNLNSKLNSEKLYIRIGIRKGGCGSGLSY FMTEKKSNTSDNDLVYDYNFTLVCCNKSILVLYGISLDYSSSLDGGFKPLPNAK QTCGCGKSGS"
ORGANISM	complement(877. . 1095)
ORGANISM	/note="similar to L28"
ORGANISM	/codon_start=1
ORGANISM	/product="50S ribosomal protein"
ORGANISM	/protein_id="AAE81680.1"
ORGANISM	/db_xref="GI:8925948"
ORGANISM	/translation="MDIKMTRKCOLLEFKTSNTANKVSHSNHKSRLQFVNLOKKKWS SLKKWKILRISTKMKILINKKNIDYFL"
ORGANISM	1332. 2813
ORGANISM	/gene="rbcL"
ORGANISM	/gene="rbcL"
ORGANISM	1332. 2813
ORGANISM	/gene="rbcL"
ORGANISM	/codon_start=1
ORGANISM	/product="ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit"
ORGANISM	/protein_id="AAE81681.1"
ORGANISM	/db_xref="GI:8925949"
ORGANISM	/translation="MSQSLEESVQERTRIKNSRYESGVIPYAKMGWPNDFVKKDID VLAEPFPOGVDPIEAAAIVAGESSTATATVWFDLPLAADLYRAKAYKVOOVNPN PBOYFAIAYELDLFEESGTANLTASTLIGVPGKAYKALRLEDMRPFAYITPGCP ATGVILIERERLDKFGRLPCLCTTKRKIGLSGKNYGRVYALKGGLDFVDDENINSQ PEMKREYRLFVEAVNKAATAAGVEVKGHLLNVAATMEBTVARAQDAKLGSVITMT DLVIGYAIQTGMKAWARDNDMLHLHRAAGNSTYRQCNHGMNFRVYCKMKMRAGVGH HAGVVGKLEGGDPIITRGFYKTLPLPLERNLDGELFEDMDASLKRKVPVASGGIHA GOMHOLIHVYGEDVLOFGGCTIGHPDGIGSGATVANAVALAEMILANRENRDPLTECP ELIREAKNGCALRTALDLMKOLITENTSTSDSVFETPIANI"
ORGANISM	2859. 3278
ORGANISM	/gene="rbcS"
ORGANISM	2862. 3278
ORGANISM	/gene="rbcS"
ORGANISM	/codon_start=1
ORGANISM	/transl_except=(pos:2862. . 2864,aa:Met)
ORGANISM	/product="ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit"
ORGANISM	/protein_id="AAE81682.1"
ORGANISM	/db_xref="GI:8925950"

CDS					/translation="MRTQGTSEFLPDLTQEOIKQIDYMSKLAIGIEYTNIDHPH NSFEMKGLPLEFTDPAVLFEIENACRAKSNFYIVGVSSRGIESTIISHIVNR PKHPGPNLRQEKSRSIKSIDAYTYTPEDORY" 3391..4308 /note="similar to CfxQ/Cfxx protein" /codon_start=1 /product="unknown" /protein_id="AAFG1683.1" /db_xref="GI:8925951" /translation="MNVITLETYEKENINENTLVNIEEYKTKIQEVLNQLDKELIG LTPVKNRKEIALALLIDKLRIKLNTQANGLMSTGSGCTQKTVAKRMADITLV LGVYRKHGLITLVREDLVGVYIGHTARCTEVLRKAGGVLYFIDEAYLYKPDNERDY GSEADIELLOVENQRNDIVLAVILAGYDKMREFOSSGSRITNVDVDPYAEEL LEIGIMLEEOQYQLTPEEGKVVILEYIKRRMTQPHFANMSIRINADIRAKNRARLIF NNSRNVITTKDVLVTIEADILKSSLEFMDK" complement(4506..5051) /gene="ahas" complement(4506..5051) /gene="ahas" complement(4506..5051) /gene="ahas" /codon_start=1 /product="acetylhydroxy-acid synthase small subunit" /protein_id="AAFG1684.1" /db_xref="GI:8925952" /translation="MKRTSLVLYDEDSGVLTIRISGLFARGFNIDSLAVGPAEKAGIS RTVWPGDNDRITQITKLHLYLVNLIKQIDINTIPCOVERIMIKYTGDKNRNELLT EIANIFRVRVDSSEFIILETIGDPKIAIEKLEKYGICEIARTGKIALTRFSKI NTELRNTMINNDKIRGGT" 5201..5746 /gene="if3" 5201..5746 /gene="if3" /codon_start=1 /transl_except=(pos:5201..5203,aa:Met) /product="translation initiation factor IF-3" /protein_id="AAFG1685.1" /db_xref="GI:8925953" /translation="MLEKSKOKNSFDKNKFLINKISFPYIRIIDENNYOLGIKIED ALKIAEKRNIDVLINDKTEPPYRIITIDYGFKAQKSKSEAKKHQVYVEIKNR YKIEDHYQVRINDAIKFLIKGNKVKISTIFEGREIYIDLAELIKIHTISLAE HEKVFDEKGEINVIYMLPKRT" 5786..6505 /note="ORF720, putative ABC transporter; Amp-binding protein" /codon_start=1 /product="unknown" /protein_id="AAFG1686.1" /db_xref="GI:8925954" /translation="MKILIEKICKFYRKCTIDNLHMTFSOYETIGLIGPGCTGTSI FKLITGISKPKDKVILNGSDITVNPVHQARIGCTCLPOBSSVFRRLTYKENIKLAL BEKYSDAIESIISTFTSYFQIDTKITHTGSYISGGERKLEIARVYIATKPKLLD EPFAGIDPISISIELQETIIVNLNNSGIGILITDHNVRETLKITRSHYILSGINILASCT TROLIENKAVRESVLGNFTL" 6553..6981 /note="ORF429" /codon_start=1 /product="unknown" /protein_id="AAFG1687.1" /db_xref="GI:8925955" /translation="MFWIIRYILQLLIGPVISSAFTSTLTSVGLFELVRRVVEG LPFKAISILLVKSPIYFIFLSPLASFLRILITPNOANKEKELIGFSLGISNKKWLIR STIRGLILAISSFINDLIVPVSNNHRANELRESKENTL"
CDS					
CDS					
gene					
CDS					
CDS					
gene					
CDS					
CDS					
BASE COUNT	2747	a	894	c	984
ORIGIN					g 2357 t
Query Match	9.4%	Score 156.6;	DB 8;	Length 6982;	
Best Local Similarity	58.5%	Pred. No. 2.3e-33;			
Matches 273;	Conservative 0;	Mismatches 194;	Indels 0;	Gaps 0;	
999	CATACCTCATCATGCTCTTAATGATATTCACAGAGTTCTTAATATGAGTGGTGT	1058			
Db	5045	CATACATTATCTGTTTGTAGAGATGAATCCGCGTGTTTAACTCGCATTTCTTGATTA	4986		
1059	TTTCGCTCGAAGGGGATACAAATATTCAGAGCTTGGCCGTAGACATGCTGAACCAAGGGC	1118			

Db	4985	TTTGCTAGACGTGGATTGAATATAGATTAGCTTAGCAGTTGGTCTGCTGCAAAAACGACGT	4926
Qy	1119	ATTTCAGCATTACACAGGTTATACCTGCAACAGATGAATCGTCAGCAATTTGGTGCAG	1178
Db	4925	ATATCAAGATTACAAATGCTTATATCAGGGGTGATATATAGAAATTATAGAACAAATTTCAAAA	4866
Qy	1179	CAACTTTCAAACACTCGTATGATGTCATAGCTCCATGATCTTACTCATTTGCCATTTTCT	1238
Db	4865	CATCTTTATTAATATAGTAAATGTAATTAATAAAATTCGAGTATATACAAAATTTCCATGTGA	4806
Qy	1239	GAAGAGAACAGTATGCTATTAAGATTGCGGTGAACGCTGCGTAGAAAGATGTCTGTG	1298
Db	4805	GAAGAGAAATTAATGCTATTAAGAGTGAACAACTGGAGAGAAATTAAGAAATGAATATTTG	4746
Qy	1299	GACATTTGCTAGTATTTTCGAGGGCTAAACGCTTTGACGATCTGATCACACATTTACTTTG	1358
Db	4745	GAATTTGCCAATATTTTATAGAGCACAGTGTGTTGATTTATCAGAAATCTTTATATTTTGA	4686
Qy	1359	CAGCTTACTGGGGATCTATGACACAGATGTTGTCACCTGCACAGGTTATTTGGAGCCGTATGT	1418
Db	4685	GAGATTAACGAGAGATCCCGCAAAATCCGCTGATTTGAAAAATTTATGAAAAAATATGT	4626
Qy	1419	ATATGTGAGTTTGCAGAACGCGTCTGTGCAATTTGCTCGTATTC	1465
Db	4625	ATTGTGAAATTTGCTAGAACGCAAGATGCTCTTACACAGAAATC	4579
RESULT 8	AF022186/c		
LOCUS	AF022186	164921 bp	DNA
DEFINITION	Cyanidium caldarium strain Rk1 chloroplast, complete genome.		
ACCESSION	AF022186	Z36235	Z70297
VERSION	AF022186.2	GI:8465296	
KEYWORDS			
SOURCE			
ORGANISM	Cyanidium caldarium.		
	Chloroplast Cyanidium caldarium		
	Eukaryote; Rhodophyta; Bangiophyceae; Porphyridiales;		
	Porphyridiaceae; Cyanidium.		
REFERENCE	1 (bases 130696 to 132364)		
AUTHORS	Vogel,H., Fischer,S. and Valentin,K.		
TITLE	A model for the evolution of the plastid sec apparatus inferred from secY gene phylogeny		
JOURNAL	Plant Mol. Biol. 32 (4), 685-692 (1996)		
MEDLINE	97134960		
PUBMED	8980520		
REFERENCE	2 (bases 1 to 164921)		
AUTHORS	Gloeckner,G., Rosenthal,A. and Valentin,K.		
TITLE	The structure and gene repertoire of an ancient red algal plastid genome		
JOURNAL	J. Mol. Evol. 51 (4), 382-390 (2000)		
MEDLINE	20496959		
PUBMED	11040290		
REFERENCE	3 (bases 46857 to 47851)		
AUTHORS	Valentin,K.		
TITLE	Direct Submission		
JOURNAL	Submitted (22-MAR-1996) Institute for Plant Physiology, Justus Liebig University, Heinrich Buff Ring 58-62, Giessen 35392, Germany		
REFERENCE	4 (bases 28701 to 75580)		
AUTHORS	Gloeckner,G., Rosenthal,A. and Valentin,K.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-SEP-1997) Department of Genome Analysis, IMB Jena, Beutenbergstr.11, Jena 07745, Germany		
REFERENCE	5 (bases 1 to 164921)		
AUTHORS	Gloeckner,G., Rosenthal,A. and Valentin,K.		
TITLE	Direct Submission		
JOURNAL	Submitted (18-NOV-1999) Genome Analysis, Institute for Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany		
REFERENCE	6 (bases 130696 to 132364)		
AUTHORS	Vogel,H., Fischer,S. and Valentin,K.		
TITLE	Direct Submission		
JOURNAL	Submitted (18-NOV-1999) Institute for Plant Physiology, Justus Liebig University, Heinrich Buff Ring 58-62, Giessen 35392, Germany		

COMMENT On or before Nov 23, 1999 this sequence version replaced gi:529651,
gi:1240002, gi:2465730.
FEATURES
source Location/Qualifiers
1..164921
/organism="Cyanidium caldarium"
/organelle="plastid:chloroplast"
/strain="RK1"
/db_xref="taxon:2771"
repeat_region join(164808..164921,1..78,462..644)
/note="direct repeat;
repeat unit separated by stem loop"
stem_loop 339..477
/note="stem-loop separates direct repeat unit"
580..780
/note="similar to yeast mitochondrial origin of
replication region in Genbank Accession Number U36902"
gene 986..2041
/gene="psbd"
986..2041
/gene="psbd"
/product="unknown"
986..2041
/gene="psbd"
/note="Photosystem II D2 protein"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAFI3026.1"
/db_xref="GI:6466445"
/translation="MTAIGREQERGFDLLDWIKDRFVIGMSGILLPCAYIAL
GAMFTGTTFVSWYTHGLASSYLEGCNFTAAVSSPANSMSGSLLEFLMGPEAQGFTR
WCOIGLMTFTALHSGFLIGFLROFEIARLVGLRPYNAIAFSGPIAVEVSFLLP
LGASWEPAPSPREVAIAIFRFLLEFLOGEHNMTINPEHMGVAGILGALLCAIHGAVE
MTLPEFGASDPFRAPTPROSEPTYSMATVNSQIRGVAPANKRMLPFLLEFVTE
GLWVSSIGIVGALNLRAIDFVSQETRAADEPEFEFTYKNIILNCGITAMMAADDP
HENVFPEEVLPRGNAL"
gene 2031..3413
/gene="psbc"
2031..3413
/gene="psbc"
/product="unknown"
2031..3413
/gene="psbc"
/note="Photosystem II 44 Kd apoprotein"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAFI3025.1"
/db_xref="GI:6466444"
/translation="MLFNKDKNNIGRSIESGTGFAMWSGNARLINSGKLGAHVAA
GLIVFTGAMTLEETSHFLPEKPLYOGMILLPHLATLGMGVAPEGGEIVNTYPYATG
VTHLVSSAVLGRGCIYHSTVGPVLEDSFSFGYDRODKNNKMTTIGIHLILGIGAP
LVYKALFTGGIYDWAPOGGDIRFTINPTLNPALITFYLSPGEGEMIVGVNME
DVIGIHWIGVTCVIGIWHILIRPFSMARAFVWSGEAVLYSLGALAMQTALEY
AWYNNVYPSSEFGYPAEAESAQAFTFLVRDQRLGANIASITOGPYGLKYLMSPTG
EVLGGETMRFMDLRAPWLEPLRSSNGLDINKINDIOPQERRAAEYVMTIAPLSIN
SVGGVATEINSVYVPSRSMLTTSHFLEFPFIIGHLMHAGRARARAAAFPEKGINEN
EPLVSMRPID"
gene 3546..3821
/gene="ycf19"
/note="similar to ycf19 in Porphyra purpurea"
3546..3821
/gene="ycf19"
/product="unknown"
3546..3821
/gene="ycf19"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAFI3024.1"
/db_xref="GI:6466443"
/translation="MKDGDILILETTIAFLQIYIVILLRMSLGWFPNINWYSPQPS
LSQSDPYINLPHGVFPSPFLGIDFSPILIGITLIDFTIELLSROLKPL"

gene 3873..4133
/gene="ccrf9"
3873..4133
/gene="ccrf9"
/product="unknown"
3873..4133
/gene="ccrf9"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAFI3023.1"
/db_xref="GI:6466442"
/translation="MNNCLEPKNKVKFELNFKCIDPMCTDNLKFPPLKLNKKGKFT
NKFYNTKNKKYIDILFIYIYHYILRIGKNQTKDLQITKC"
gene 4189..4272
/gene="ccrf10"
4189..4272
/gene="ccrf10"
/product="unknown"
4189..4272
/product="unknown"
4189..4272
/gene="ccrf10"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAFI3022.1"
/db_xref="GI:6466441"
/translation="MTKLKSLNDKKLLISCLNKTIVDIN"
4334..4406
/product="tRNA-Thr"
4450..4971
/gene="ycf3"
/note="similar to ycf3 in Porphyra purpurea"
4450..4971
/gene="ycf3"
/product="unknown"
4450..4971
/gene="ycf3"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAFI3021.1"
/db_xref="GI:6466440"
/translation="MPSQKNDNFIDKTFSSIADILVKILPTNKSKEAFYXXKDMA
AQSEGEYALACYYOALKIERKDPMDKSFILYINIGLQASNGQHARALEYHESLKF
PNLYOALNNIAYIYHYGNKLPQOSKLOBAKIMPDKASWYWKAKIPLAVNYEAMNW
LKITGRIPEDIML"
gene 5004..6476
/gene="atpB"
5004..6476
/gene="atpB"
/product="unknown"
5004..6476
/gene="atpB"
/note="ATP synthase CFI beta chain"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAFI3020.1"
/db_xref="GI:6466439"
/translation="MKPCQLKOKORTLAIAKYNNKKYIQTIGPVLDIVFPGDGLPKV
FNAIKINNSNNMNTICEVOQLGDNKVRVAVASTEGRLGASALIDTGEPIISPKE
TLGRIFENVLGEPIDEKGPVISNDKLPIDRPAPKFTOLETKPSIFETGIKVPDLAPYR
RGKIGLFGGAGVGVTLIMELINNVAKAHGVSVGGVGEPTREGNDLYOBKRESGV
INBKDNLKVALCYGQNNPPGARRRVGLTALMAEYFRDYNKQNVLLFDINIRFV
QAGSEVSALDGRMPSAVGYOPTLGTBMGALQERITSLDGSITSIQAVVVPADDLTDP
APATVFADLADATVLSRALAAGIYPADVPLDSTSTMLOPGIVSDEHVTYAKVETL
QRYKELODIIAILGDELSEEDRLIYSRAARKTEKELISOPFVAEVEFTGISGRYVSLSD
SIKGFNMILSGEVDNIPQOAFYLVGRIEBAIDKAKOVERS"
gene 6486..6887
/gene="atpE"
6486..6887
/gene="atpE"
/product="unknown"

CDS
6486..6887
/gene="atpE"
/note="ATP synthase CFI epsilon chain"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAFI3019.1"
/db_xref="GI:6466438"
/translation="MALTVKVTTPDRVYKKTVETIILPSSNGSLIMNNAPIILTL
DIGVRAVMVTPVLYLGLGFAQIDNNLVITISDAEYKAIIDEPEANKLIASLAN
MEKQSNREKIESMONLRARRRMOALISLK"

Query Match 9.2%; Score 153.8; DB:8; Length 164921;
Best Local Similarity 59.7%; Pred. No. 2.2e-32;
Matches 279; Conservative 0; Mismatches 182; Indels 6; Gaps 1;

QY 294 AGAAGCACACATTTCAATTTGTTGGAGACGAAGCGGAATGATTAATAGATTGCA 353
DB 106832 ATGAATATACCTCTCCCTATTAGTGAAGATGAAGCTGGCTTTGACACGTATCGCA 106773
QY 354 GGAGTGTTCAGAGAGATTAATTTGAGAGCTGCTGTTGG-----TCGAGC 407
DB 106772 GGGTGTTCGCTAGACGAGAGATTTAAACATAGAAAGTTTAGCTGTGGACCTGCTGACAA 106713
QY 408 AGAGACAGGCTCTATTCACATAGTTGCTGTGAACTGAAAGGGTACTTCAGCAGTTC 467
DB 106712 AAAGGATCTCAAGATTCACATAGTTGTTACCGGTGATGATCGCAGCTGACAGTTA 106653
QY 468 ATCGAGCAACTCCAGAGCTCGTTAATGTTCTAAGGTTGAGATATCTCAGTGAACCG 527
DB 106652 ACAAAGCAACTTATTAAGATTAATTTTAAAGATTAATATTCAGAGATTCCCT 106593
QY 528 CAAGTGACGCTGAGCTGATGCTGTAAAGATGAATGACATCCAGACGAGGAGAG 587
DB 106592 TGTGTTGACGCTGACCTACTACTAGTAAATTAATGCTAGTACTGTTCCAGACAGAA 106533
QY 588 ATCATGTGCTAGTTGACATTCAGACAGAGTTGTAGATATAGCGGAACATGCATTG 647
DB 106532 ATTTTAAATTTTGGCAATTTTGAAGCTAAAGTAGATTATTCAGAAATCTGTTA 106473
QY 648 ACTATCGAGTACCTGAGATCTCTGGAATAATGATTGCTGTGAAAGAAATTTGAAAAG 707
DB 106472 GTTTTAGAATTCACAGGAGTCCAGGTAAGTAGCTCAATTCACAGCTACTGCAAAA 106413
QY 708 TTTCAGATCAGAGATTTGTAAGGACAGGAAGATAGCACTGAGAG 754
DB 106412 TTTTGAATATATAGAAATTCAGAGAACTGGCAAAATAGCGTTAACAG 106366

RESULT 9
AP005371 300450 bp DNA linear BCT 17-AUG-2002
LOCUS Thermosynechococcus elongatus BP-1 DNA, complete genome, section 3/9
ACCESSION AP005371 BA000039
VERSION AP005371.1 GI:22294315
KEYWORDS
SOURCE Thermosynechococcus elongatus BP-1 (strain:BP-1) DNA.
ORGANISM Thermosynechococcus elongatus BP-1
REFERENCE 1
AUTHORS Nakamura,Y., Kaneko,T., Sato,S., Ikeuchi,M., Katoh,H., Sasamoto,S., Watanabe,A., Iriuchio,M., Kawashima,K., Kimura,T., Kishida,Y., Kiyokawa,C., Kohara,M., Matsunoto,M., Matsuno,A., Nakazaki,N., Shimo,S., Sugimoto,M., Takeuchi,C., Yamada,M. and Tabata,S.
TITLE Complete genome structure of the thermophilic cyanobacterium Thermosynechococcus elongatus BP-1
JOURNAL DNA Res. (2002) In press
REFERENCE 2 (bases 1 to 300450)
AUTHORS Kaneko,T.
TITLE Direct Submission
JOURNAL Submitted (05-JUN-2002) Takakazu Kaneko, Kazusa DNA Research Institute, The First Laboratory for Plant Gene Research; 2-6-7

Kazusa-Kametarai, Kisarazu, Chiba 292-0812, Japan
(E-mail:kaneko@kazusa.or.jp)
URL:http://www.kazusa.or.jp/cyano/Thermo/
Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)
location/Qualifiers

FEATURES
source

1..300450
/organism="Thermosynechococcus elongatus BP-1"
/strain="BP-1"
/db_xref="taxon:197221"
/note="BP-1"
126..1655
/gene="c1r0594"
126..1655
/gene="c1r0594"
/note="c1r0594"
/note="ORF_ID:c1r0594
competence protein ComM homolog"
/transl_table=11
/codon_start=1
/protein_id="BAC08146.1"
/db_xref="GI:22294316"
/translation="MLARWSSAAYVIGIDAIPIYGVVDYSGGIPRIYVYVGLPAGYQEA
HEVRALRNAGSFSPGQILVNLTPADLRKGFSPDLPISTIGILASGQVATDLIGD
HLFLGLSDTGLQPVAGVLAIALAAGQITGLVVPANAMATEALVGLKYGGCHTL
AEVVAFLNDPSSRSPATPSPCWAAASPOFTVDLKDVGQYQARALEIAAGGHNLI
FVGPSSGKTMARLPILPLTPFAELEVTKHSVAGLIDKERGLQDEPFERSPH
SASGAIVGGSGYPRRGESLARGVLFDELTERKRDYELLRQPLDGQVTSRAR
QSVFPAQFTLVASTNPPCGTYGPGVQPTCTSPKROQYAKLSGLPLDRIDLOVY
SLKPEEMTRQPLGEDSATVRQVLAASRAQORFAEENVHNCNOMQSHLRQWCSL
DASIHLELRAIQGLSRAVTRDILKVARITADLADDETITSAHVAEAIOYRTIDRL
Q"
1588..2898
/gene="c1r0595"
1588..2898
/gene="c1r0595"
/note="c1r0595"
/note="ORF_ID:c1r0595
probable sugar transporter sugar binding protein"
/codon_start=1
/transl_table=11
/protein_id="BAC08147.1"
/db_xref="GI:22294317"
/translation="MKOLPLPMWRKRPNTAQLTYCNRGWFLILGLFALVIGCTTA
TQSEQPIELMHWGYNPPRNRYLQRLVDRFNARHQLHVQALVYGGPDDOLPLIAA
VGNANPDLVLPNTITGQFVELGLAPLDDMOWASPYRDQSPALITPMRGNYHS
IPFANNLGIYRPSLTPFAAGIKTLPSMQLLEQVYSQTLKAGITPILPIGGGFAY
FSLDFPNSAGGSISCTEPETAHITDPAMALEFQRLRQRLAFLSPERGETLDOF
LRGEVAMQITGPWTGLQDQSGVDGVLPMPLKTPATALGGENLFVRSPTQREAA
LAFILNVLSSEFQTEWALGTGLPINERVTRNDRYRAFEVAQDPLRFLVDMAAQAAR
PNFRGVARFSQTLGRALLESVILGKSTPKAAVYTTABRQWLMRPR"
complement(2895..3905)
/gene="c1l0596"
complement(2895..3905)
/gene="c1l0596"
/note="c1l0596"
/note="ORF_ID:c1l0596
probable heme d1 biosynthesis protein HtrJ"
/codon_start=1
/transl_table=11
/protein_id="BAC08148.1"
/db_xref="GI:22294318"
/translation="MGVHIKQAVEVATYIISQRLQGRKRFPLTLMLEPLFRCNLACSG
CGKIQHLEVLRLKYLTPCECFRVEECGAPIVAIIGGEPILHPOIDEIVSGLVARRK
VYLCTNGILLENLHKFPKSPYFSPVHLDGLREWHDCVARKGVEDVAIVAKAKA
KAFRTYNTITTEGYDPEKIEQFPFQVSLCYDGMISPGYAVBNAPQDHFKREOT
RALFEITLAPTRAGCKRMFNHNPILDFLVEKYECTPGWMPSTSVLQWQPCYLL
NEGHRTYQWELLENDMDKYGKHSNPKCRDCMVHCGYEATRAMDAFPMANVKAASS
LF"
4441..6447
/gene="c1r0597"
4441..6447
/gene="c1r0597"
/note="c1r0597"
/note="ORF_ID:c1r0597"
/codon_start=1
/transl_table=11
/product="ribonuclease E"

gene
CDS

```
/protein_id="BAC08149.1"
/db_xref="GI:22294319"
/translation="MPKOIIIAEHRILAEVAEDQIOLIVATGTHQVDIYLGVEN
VLPGLDAEAVNLGDASRNGFIHVDLGLRLKRTGATLTLLAPOKQVIVOMKPPG
TKPRLTGNISLPGRYVILMPYGRGVNLSRRTNAEHRILRAIGLTKPMGMLIVR
TEAEVSEALIEDELLQROMETIQOQAASRPOLLGRDDFIQKVLROYSSDVN
RIVVTPDGKVRKQHKLMNVNRPVGLIDHOEPIITLDFRYNAIIRELKRVD
LPSGGYIIIOPTREALTVIDVNSGFTSSATSHETVLTWNCENATEIARQLRNTAGY
IIVDEIDMDSRDDQLLELHEFSKALRADKARPOIALSELGLVELTRKROGNIYELF
GRPCACGGGLGHLVHLPGEDESPGEIVERSPIPRPPRSLERTEPENGWASSLE
TOLINHPYOAVGIRPRSRSSRSHGVPTRSPRPPENGKRVVTTTPEEAAPPVYVR
ESRYPYRPARPSSEPAVITYTMTDOEAVYAEMGISPLVLXGEEVKNPRTGIYVAV
EGDDPLPQKTRGPVSEGGPEAPEIADLSPSEVSTWNEPIPATGLTLNTSESLKT
SEPEVTAEASGEASTPEASVPIGRRRRRVTSSEQLSLDNA"
/gene="L1r0598"
/length=7340
/size=6456..7340
/gene="L1r0598"
/note="ORF_ID:L1r0598
putative glutaminase"
/codon_start=1
/transl_table=11
/protein_id="BAC08150.1"
/db_xref="GI:22294320"
/translation="MLGQHPDLIOEWLEAATEIPRGEPIQLRPKVSSEAFEAVALT
PEOEFILGRSLRFLPMSAIKFFLLVALTLMOEQVMTWVGORPDPYNSVLTQLTLD
QOEPNPMINSGLALAGOLAVYGATFOITWNECAGTOFEVDOEVLGAVHRHNMQ
NNTLAFIVESGRIDAEVALEITYNQCCFOGTVEEVARLGLLALPOAKVSDRHROO
NVNMLTGCIVEDSPRYGLEIGLPMKSGVGMALIVRQCAIACYSPPLDKSGNSIL
GLYLQIRISRHGLSPLC"
/complement(7287..7823)
/gene="L1l0599"
/complement(7287..7823)
/gene="L1l0599"
/note="ORF_ID:L1l0599
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAC08151.1"
/db_xref="GI:22294321"
/translation="MTADPIAGCAMLROALALAGFKTTAVAVESPLAGVVGHWLEI
DROSLSIAOIESLANMNOALDLOYLLNATLNGCRSREBOATFLELAGIRAKYRQ
LOKIAHRAVARLDGKPELEIRLSAERRVHSLIRLESRLTFESHGEEPRRLV
DLAQTEAEVTDALDEQY"
/complement(7832..9037)
/gene="L1l0600"
/complement(7832..9037)
/gene="L1l0600"
/note="ORF_ID:L1l0600
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAC08152.1"
/db_xref="GI:22294322"
/translation="WDFGIGFLSNVMWLPIIDFYGIVPSGLAIVALTLVREAVYP
LSAGSIKMMRRKVVOPIMOKRMOEIQKYNKDNPEQOKAAAEVAREGNPLAGCEPL
LIQPLILFALFATLGRSPADVNOINLOIYPPDPVALQEPATPOMIYVDEGHAF
IOAVVSGTRIPVGMVDLOPOTKEGKPESELVAVYPSHLOPKMOIIRGOEARIDR
EGHLOALDGEVITGCTIGIAADKGLFISLGHVGIIDGCELTITPDGARONMG
AIHMDILVMIIGFVSLYVNOILSGOGSSNNQOOTVNOITPITISGAFLEPPLAGV
LIVLIANIYFOTVOTFILOREPLPENLOKLVEEDERAIRIKNRAPLPEGGAKGAKA
TKAKGNPSS"
/complement(9085..9822)
/gene="L1l0601"
/complement(9085..9822)
/gene="L1l0601"
/note="ORF_ID:L1l0601
probable ribonuclease P"
/codon_start=1
/transl_table=11
/protein_id="BAC08153.1"
/db_xref="GI:22294323"
/translation="MLPROHRLRDRNDEDRVYQHKTAHHGLAVLMLAPOSSEAPRL
```

```
AIIVSKVSKRAVDNRNLKROWRYRIFLPLQMAAGMDIIVYRQALHVSDEDCRTE
LTQLMAKAGVLTMPGISSEVYEGGPHGDLVINDLGLTYICLPLTGAIVRALMLR
YRITRRTITVIGWGMGSDSDYIEIRIKVTVTPRGWAGYDMVVTLDGSRLEMRAY
PFRFRIYIYISFKLSLPAKSVSGAIGS"
/complement(9822..9959)
/gene="L1p134"
/complement(9822..9959)
/gene="L1p134"
/note="ORF_ID:L1p134"
/codon_start=1
/transl_table=11
/product="50S ribosomal protein L34"
/protein_id="BAC08154.1"
/db_xref="GI:22294324"
/translation="WTQRTLGTVAKRKRTSGFARMRTKSGQNVIKARRRGRARLT
V"
/length=10140..11501
/gene="dnaA"
/length=10140..11501
/gene="dnaA"
/note="ORF_ID:L1r0603"
/codon_start=1
/transl_table=11
/product="chromosomal replication initiator protein"
/protein_id="BAC08155.1"
/db_xref="GI:22294325"
/translation="MISAEEDLMHOILERLLOLLSRPFTETWIKATVQSPFGOTLTI
CPANFARNLOKYYLKITADYARIVAKPEITELAIQAEHDSIRPSTRGEVAT
PPAARHRSSELNPKYFERYVYVGRPNRMAHACLAIVASPERENPLFLCGVGLGT
HLMQAIHGRLEIDPNAKITVYSTEQFINDLIARIKOSWSFRRHTAVYVMLVDDI
QFIEKREYTOEEFFHTFTLHAGQOVVALDRPSQIPRLOERLCSRSFSGLLADIQ
PPDLTBRMIILOKKAKEYENIRLPREVITYIASYTSNIREGLAIRAVAYISISGLP
```

Query Match
Best Local Similarity 59.5%; Pred No. 1.7e-31;
Matches 276; Conservative 0; Mismatches 182; Indels 6; Gaps 1;

QY	291	GTGAGAACACACAAATTTACGATTTTGTGGAGACGAAGCGAATGATTAATAGATT	350
DB	297278	GTTATGAACATACACCTATACGTTTGTGGAGAAATGAGCGGTTGTCTACATGCATT	297219
QY	351	GCAGCATGTTTTCAGAGAGAGATACAAATTTAGAGCTTGTCTGTGCTCT-----G	404
DB	297218	GCGGCTATTTTGGCGGCGCTTTAACATCGAGAGCTTGCGGTAGACCTGCGGAG	297159
QY	405	AACAGACAAAGCTCTATTACCATGATGTTGCTGTGGAAGTGAAGTACTTACGAG	464
DB	297158	CAATTTGGCATCTCCCGTATTACGATGTGTGCTGAGATGATGCTCTATTGACGAG	297099
QY	465	TTGATCGAGCACTCCAGAGCTCGTTAAATGTTTGAAGTGAAGATATCTCAAGTGAG	524
DB	297098	TTGACCAACCACTTCAGAACGATCATGTCCTCAAGGTGAGAGATVTTACACAAAT	297039
QY	525	CCGCAAGTGGAGCTGAGTGTGTTGTTAAAGATGATGCACATCCACAAATCCAGGCA	584
DB	297038	CCCTGTGTGAACCGGAGCTGATGCTCCGAAGGTGAAGTCCACATCCGACCCGCTCT	296979
QY	585	GGATCATGCTGGCTAGTACACATTCAGAGCAAGGTTGATATATACCGGAACATGCA	644
DB	296978	GAAATTTGGAATTTGTCGAATTTTCCGCGCTAAGGTGTGATGTGTCGACGATTCG	296919
QY	645	TTGATCATGAGTAACTGAGATCTCGGAAAAATGATGTTGTAAGAAATTTGAAA	704
DB	296918	CTCATCATGAGAGTCTCCGGTGAACCTCGCAAAATGTTGCTTGTGCAAAATGCGAAT	296859
QY	705	AAGTTTCAGATCAGAGATTGTGAAGCAGAGAAGATATGCACT	748
DB	296858	AAATTTGGATTCCGGAATTTGCCGACCGGTAAATATGCCCT	296815

RESULT 10
AE013291/c
LOCUS AE013291 11318 bp DNA linear BCT 17-MAY-2002

DEFINITION	Methanosarcina mazei strain Goel, section 73 of 379 of the complete genome.
ACCESSION	AE013291
VERSION	AE013291.1
KEYWORDS	GI:20905057
SOURCE	
ORGANISM	Methanosarcina mazei Goel. Methanosarcina mazei Goel. Archaea; Euryarchaeota; Methanococci; Methanosarcinales; Methanosarcinaceae; Methanosarcina.
REFERENCE	1 (bases 1 to 11318) Depemeler, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R.A., Martinez-Arias, R., Henne, A., Wietzer, A., Baumeier, S., Jacob, C., Brueggemann, H., Lienard, T., Christmann, A., Boemecke, M., Steckel, S., Bhattacharyya, A., Lykidis, A., Overbeek, R., Klenk, H.-P., Gunsalus, R.P., Fritze, H.-J. and Gottschalk, G. The genome of Methanosarcina mazei: Evidence for lateral gene transfer between Bacteria and Archaea J. Mol. Microbiol. Biotechnol. (2002) In press
TITLE	2 (bases 1 to 11318) Depemeler, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R.A., Martinez-Arias, R., Henne, A., Wietzer, A., Baumeier, S., Jacob, C., Brueggemann, H., Lienard, T., Christmann, A., Boemecke, M., Steckel, S., Bhattacharyya, A., Lykidis, A., Overbeek, R., Klenk, H.-P., Gunsalus, R.P., Fritze, H.-J. and Gottschalk, G. Direct Submission Submitted (23-AUG-2001) Goettingen Genomics Laboratory, Institute for Microbiology and Genetics, Grisebachstrasse 8, Goettingen 37077, Germany
TITLE	Location/Qualifiers
JOURNAL	1. .11318 /organism="Methanosarcina mazei Goel"
REFERENCE	/strain="Goel"
AUTHORS	/db_xref="taxon:192952"
	931. .1776 /gene="MM0663"
	931. .1776 /gene="MM0663"
	/codon_start=1 /transl_table=11 /product="putative dihydroorotate dehydrogenase electron transfer subunit"
	/protein_id="AAM30359.1"
	/db_xref="GI:20905058"
	/translation="MVYRILEKEELIAPSVHRMVIDAPVAKAAGOFILRIDERGE RPLFIADPEKGTGTVTFVIFOMGKTKTOLAKLSAGDSLEDVPGIGPADIRKIGTV ILVGGVAVAPVYPOAKAYVANKVVISILGARMKDILIDPEMKASSELYIANDDG SKGHGVTIDIKRILIDSGEKARVYITGPPIIMKVGAGVAPTYVEILLVNSTAVD GIMCGGCRVTVGGTFTKFTCVDPGFDAKVDVQLMNLAMYRGEETATEKYEKC RCGLH"
gene	1795. .3204 /gene="MM0664"
CDS	1795. .3204 /gene="MM0664"
	/EC_number="1.4.1.13"
	/codon_start=1 /transl_table=11 /product="glutamate synthase [NADPH]"
	/protein_id="AAM30360.1"
	/db_xref="GI:20905059"
	/translation="WQELHGEKTYGAKKERPPMPDOPAGEERRKNFNEVALGYTKDAL AASRLACKBEKCVGCPVNDIPGFIOLICEENFEGAIERIKATNALPAICGVCP OESQEAHCIVLCKGKGPVAGLERFCADYEREKVSPOVMPKGRVAVVGGSPAG LTAADDLAKIGHKVTIFESLHAKGVLSTGIDPEFLPKETIVROLETYEIKIGVEVRPN YIIGIKTIDELCDPDAVFLCTGAGLSFPGICEBENLNGVYSANPELTRVLMKAYD PEYDITVRGKNVYVGGVNAVDAKRSALRGAEVSLIVRGDEMPAREEIEHA KEEGITFLRLINPVRLIDDEKRNVTAGVCEIRNLEDPKSGRRSPVPEGSEFLVAD VVVIAIGSPNPIIFKSGEGLEOMKRGTVVADDEGTGATSKCGVFAGGVVTAATVIS AMGAGKAAKAIDEYLKER"
gene	complement(3320. .4189) /gene="MM0665"
CDS	complement(3320. .4189) /gene="MM0665"
	/codon_start=1
	/transl_table=11 /product="hypothetical protein"
	/protein_id="AAM30361.1"
	/db_xref="GI:20905060"
	/translation="MEKNIFANLGSKYSADAEITWTSASLINSATETRIKSVQDYESS TYSNAKINLNASETDPEELKILNNASSDYEEIODITFYKTLAAGLDVRHSLSLI TAMEHSDKSMAYAVSKERFNLKRKLINLNEALNESASISAKERVPFIDPKSVYIEQ KSSILLRNDELEASETMSSELROMSGMPPYMDGYVCLSNGLIETGDAEEMGKADEFG KASDKFESOKILETLTKDSEYSESVTAIEICGILTQAKDPLHIEACGRVNEKERY QANAFPNVSVYV"
trNA	complement(4987. .5058) /product="trNA-Arg"
gene	5520. .5960 /gene="MM0666"
CDS	5520. .5960 /gene="MM0666"
	/codon_start=1 /transl_table=11 /product="putative nickel-responsive regulator Nlkr"
	/protein_id="AAM30362.1"
	/db_xref="GI:20905061"
	/translation="METELRIGVSLPDTLCKPEDEIEKRGYSRSEAIRDSISYI STNEMWGNKIGHVGTIVITDHTORGSLNVLTDQHHSHLIRSHIHYDREDCEFE IVLLEGKEIIEILAEALMAIKGVRLSKLITTESNETDEPSIV"
gene	complement(6758. .7300) /gene="MM0667"
CDS	complement(6758. .7300) /gene="MM0667"
	/codon_start=1 /transl_table=11 /product="Nitroreductase family protein"
	/protein_id="AAM30363.1"
	/db_xref="GI:20905062"
	/translation="MPASGENKVIDIILNRSVREFNDPRISKEELINKLNAGRAP GLNMQPWFIVIRDEPTLIKISECHYSIIIEGAPLLAALFDTENSNRRKQVDAIG ASIQMLLACDGLGAVWLGEILKONKRVNSILHCPPELDMAVLAIGEPAPGRKTS SRKPLSELVFNKYGOKWE"
gene	complement(7560. .8567) /gene="MM0668"
CDS	complement(7560. .8567) /gene="MM0668"
	/EC_number="1.1.1.86"
	/codon_start=1 /transl_table=11 /product="ketol-acid reductoisomerase"
	/protein_id="AAM30364.1"
	/db_xref="GI:20905063"
	/translation="MAIIIVDDETTFDALKRTIAVMGYSGQGHANLHESGLNVI VGLROGSSMAKAEISDGLKVMTVDDASRAADVIMILDPKQAAVYSQIEPNLKAGD ALVFAHGENIHHNOIVPKDVPVAVAPRGPHIRYRTYEGIGVGLIAYVODATGK ARDLALSTAKGIGATRAGVETTFREPERETDIFGQVYDLCGLSALITAPVLYEAG YQPEAVFETCHEVKLIVDLITEGGLERMHVSNVTAEGGAGTVGPRVINEYSRAMEY EALERIQNGEFAKEFVLGSMVNHVPLKAMEROKEDHOLEVVGKQIRANIPWLNRKIDD D"
gene	complement(8961. .9446) /gene="MM0669"
CDS	complement(8961. .9446) /gene="MM0669"
	/EC_number="4.1.3.18"
	/codon_start=1 /transl_table=11 /product="Acetolactate synthase small subunit"
	/protein_id="AAM30365.1"
	/db_xref="GI:20905064"
	/translation="MKHFLAVLVENKSGVLSRVASLFSRGRYNIDSLAVGYTEDDIS RLTIIVHGDDHVLQVTKLNDIVIKVSDIGSDDAVERELALIKVADVDTRAELI QIANIFRAIYDVAKSMIVETVGDGKIQATLEKLLRQGIKEMVRTGILAVRPPKK V"
gene	complement(9443. .11137) /gene="MM0670"
CDS	complement(9443. .11137) /gene="MM0670"
	/EC_number="4.1.3.18"

[illegible]

TITLE	Direct Submission
JOURNAL	Submitted (17-OCT-1995) Michael E. Reich, Marine Biology Section, NRC Institute for Marine Biosciences, 1411 Oxford Street, Halifax, Nova Scotia B3H 3Z1, Canada
FEATURES	Location/Qualifiers
source	1. 191028
gene	/organism="Porphyra purpurea" /organelle="plastid:chloroplast" /strain="Avonport" /db_xref="taxon:2787"
CDS	181..1053 /gene="chl" 181..1053 /gene="chl" /codon_start=1 /product="protochlorophyllide reductase iron-sulfur ATP-binding protein" /protein_id="AAC08073.1" /db_xref="GI:1276653"
gene	/translation="MKLAVYGGKGGIGKSTTSCINVALSKRGKKVLQIGCDPKHDSF TLGELPIPTIDTLQADHYEDWPEBDVYIKGYGVDCVAGGPPAGACGGVGE TVKLKELNMFDEQADIDLPVLGVVYGGFAAPLNTADYCLITDNGFDALFANRIA ASVEKARTSLRLAGLVGRNTRDRIIDYIDCVPMPLVEVLIPIEDIRVSRKGT LFEAESDTNLAVYCDVYYLNLRDIPRPGVYKEADNELSLSDFLFNKSNNT KSTADEELDMKY"
CDS	1095..2402 /gene="chlN" 1095..2402 /gene="chlN" /codon_start=1 /product="protochlorophyllide reductase chlN chlN" /protein_id="AAC08074.1" /db_xref="GI:1276654"
gene	/translation="MSIVSDALFECEGTNGYHFCPIGCVSMLYOKIEDSFVIGT KTCGFLEONAGVMIFAEPRYAMAELEGGDISATLNDYGLRLCICQIKDRDPSVFE WICGCTTEITKMDLEGIAPKLEAETRPVTVYARANGDIYAFQGEDVFLAAMQRCFS HHYIENTNPVPIVLFESLPDPVTVQVMELKQGIIVSGMLPSKRYTELPIQEGYV VAGNVPFLSRATITLMRRRTKTLIGAFEPFGDPGTAMIEKICSVSFIKDEGLAREK DIMSLIEDYALIRGKSVFPMGDNLEVSILARFIRCGMTVYEIGIYMDKRYOAEEL ALKKSCDDMMVMPITVEKPDNYNOLDRIIDLKPDILVITGMANPLLEARGINTKMS VEPFAOIHGFTNMDLLELYTRPLRNLISLSEGMPIYSKG"
CDS	complement(2454..3929) /gene="ORF491" complement(2454..3929) /gene="ORF491" /codon_start=1 /protein_id="AAC08075.1" /db_xref="GI:1276655"
gene	/translation="MNFIDQLRLKSRPIILINTREERLEYIKNOLFCDNENKV LYCMFVDGYTSNPNDNGYAKRNPLLALELEKTPYLNLFILKDPFSFNDGLIKK LRNLSOIFKOSKNIIIVSCVNIPTLSDVITVDPLPLSKIKETIRSLTALIC VLDEILNNLNKASCOGLSVERIKRTITKTAOYGDLDSRLPIIIEKRLOINOTRL EFYIYKRNATEDIGGLDLKLSMKRRSRFSKQSPNYGIPPTKGLLVIGTGTGSLTA KAINDDTLPLRLIDGLKLFGLGVESESRRREWYTAEGLSPCVLWIDEDAFSNL YSODDSQTSARVFGFTTWLSEKTPVAVATANTIONLPSEMLRKGRFMEIFLIDP NNERLELFIHLIRIPKSWONYNIKOLSLCKKFGAELTEQAIIVESMHTASERE FSTEDIKALEOFPLAVTDKEQVESIQAMAGNRRARASL"
CDS	complement(4047..4187) /gene="rpl34" complement(4047..4187) /gene="rpl34" /codon_start=1 /product="50S ribosomal protein L34" /protein_id="AAC08076.1" /db_xref="GI:1276656"
gene	/translation="MTKGTLOGSKRRKIRISGFRRMKTPSGRSLNERRRKGKTI AS" 4333..4854 /gene="ycf37" 4333..4854 /gene="ycf37" /note="hypothetical chloroplast ORF 37." /codon_start=1

```
gene /protein_id="AAC08077.1"
      /db_xref="GI:1276657"
      /translation="MLVLPFIYLSILFLFLFSLWLSQOLKTIPLFESOPQYFLDK
      SKNDELGIEDGAFSKVCIAKYPTRAIIESHLVVKKNLSLEPENAAIIATLYNMG
      FIYYKADKKLAKNPERAIEVDGNTYVALNMLAKIYETPKTKILKAHALYDVNLIAK
      SKNIATHKKFLTK"
      /complement(4889..4962)
      /gene="trnp"
      /complement(4889..4962)
      /product="trnp"
      /gene="trnp"
      /product="trna-Pro"
      /complement(5020..5544)
      /gene="ORF174"
      /complement(5020..5544)
      /gene="ORF174"
      /codon_start=1
      /protein_id="AAC08078.1"
      /db_xref="GI:1276658"
      /translation="MIFMKTFFQNSTVGESYENNEFKFLFLDKTPEDKIELKDVYLSNT
      KININYLEQLDSDVGWVRPLKVKVIALKHSIIISLIQKNDSSRLVGFAPATSDN
      GGNATIMDVYIHDPDQGLGLGKVMHQLKQLRQAEISITITLFAEPDVISFYRKLGFI
      KPDGCKGMFWYPR"
      /complement(5807..6364)
      /gene="psaf"
      /complement(5807..6364)
      /gene="psaf"
      /codon_start=1
      /translation="photosystem I reaction centre subunit III"
      /product="photosystem I reaction centre subunit III"
      /protein_id="AAC08079.1"
      /db_xref="GI:1276659"
      /translation="MKKATFITCLLAVLVSNPIYNAEVLGILPCKDSAAFRKRMVN
      SKKLQARLAKYDANTPALANKOLEKTKTPEATYGRAGLLCGTGLPHLISDGWS
      VAGDFPEGLFLYITIGWIGWGRGLSVATSKRTEKEIILDPVLAVKFPSSSGRAW
      PLAAQFESSGQLIAPNDITVSPR"
      /complement(6377..6505)
      /gene="psaj"
      /complement(6377..6505)
      /gene="psaj"
      /codon_start=1
      /product="photosystem I reaction centre subunit IX"
      /protein_id="AAC08080.1"
      /db_xref="GI:1276660"
      /translation="MNNFTKYLSTAPVIGVLWMPPTAGFIIELNFPFDVLYFYL"
      /complement(6660..7145)
      /gene="apcd"
      /complement(6660..7145)
      /gene="apcd"
      /codon_start=1
      /product="allophycocyanin gamma chain"
      /protein_id="AAC08081.1"
      /db_xref="GI:1276661"
      /translation="MSLVSQLIIMADDELRYPTIGELQSIQNYLTGSMRIRIATITIR
      DEKETIIQASAKOIFQIAPEYIAPGNAAGSKRSCLRDYGVYLLIYGVLAGKRD
      SIETIGIIGVREMYNSLGVPIIGMDIAIOCLKEASLEMLGODDAKIISPFYIIIRGM
      S"
      /complement(7473..8453)
      /gene="fabH"
      /complement(7473..8453)
      /gene="fabH"
      /codon_start=1
      /product="beta-ketoacyl-acyl carrier protein synthase III"
      /protein_id="AAC08082.1"
      /db_xref="GI:1276662"
      /translation="MGVHILSTGSSYPNFSVENQOFEDIIETSDHISTRTGKKRHL
      ABSSTSLIKLAEAANDALSKASINEDIIDLILATSTPDULFGSASQLQAEIGARIS
      TAFDITAAACSGFIIALVTASQFIQAAGSYKRVLVGADTWSRMIDWSDRSSCLIFGGA
      GAVLIGESSINSILGFKLCQGLNSHLQDLMSPSDQCGFLTTPVPGKGRSDIRMGK
      EYKFAFQVPIYIKNCILNDVNSIDEVDMFLIHKANIRILEAIATRLSIPLSKMITN
      LNYGNTSAASIPALDEFAIKEKKIOPGVVYVLAGAGAGLTGCAITVAKNQ"
      /complement(8465..8536)
      /product="trna-Leu"
      /complement(8735..8854)
      /gene="trna-Leu"
```

```
CDS /gene="psbx"
      /db_xref="GI:1276663"
      /complement(8735..8854)
      /gene="psbx"
      /codon_start=1
      /product="photosystem II protein X (4.1 kD)"
      /protein_id="AAC08083.1"
      /db_xref="GI:1276663"
      /translation="MTPSLSFLNSILGAVIIVPITLALLFVSQKDRTRS"
      /complement(9115..9981)
      /gene="psbv"
      /complement(9115..9981)
      /gene="psbv"
      /codon_start=1
      /product="acetyl-CoA carboxylase carboxyltransferase beta
      subunit"
      /protein_id="AAC08084.1"
      /db_xref="GI:1276664"
      /translation="MSLSNWPPLKREDSKRNHISTKQITIPDGLWKCFCGLMYSK
      VLKRNILKVCPOCSYHFOASSNERIDQLIDOGSMEPMVDVILISTDPLGFEDKLYSQR
      KPLAFKTSLODAVQGVGTMLGOKVCLGIMDRPFGMSGVGKLTPLLEKATQER
      LPAIILCASGAKMDEGMTSLIMOMAKTSSALFMHKKENILYSVLTSPPTGCVTASFA
      MGLDLIAEPKALIAFAGRRVIEQTIKEDLPNFOSSSEITFEHGFLLIVPTQLRSK
      LIQILHLHTNNN"
      /complement(10259..10750)
      /gene="psbv"
      /complement(10259..10750)
      /gene="psbv"
      /codon_start=1
      /product="cytochrome c550 (oxygen-evolving complex
      component)"
      /protein_id="AAC08085.1"
      /db_xref="GI:1276665"
      /translation="MLKRSSWPLATILIGLTVASVSYIYVAIEUDEATRTVPLESGR
      TVLTPEQVQRGRLEPNNSCAICHNGISITRTNRPNGDPESLIGATPQINIEALNDYM
      KQPTSIDGAESTIELHPSTIKSAEIFPKMRNLDEDLITAGHLLQPKIVSEKWSGK
      IY"
      /complement(10915..11247)
      /gene="petJ"
      /complement(10915..11247)
      /gene="petJ"

Query Match      8.8%; Score 147.8; DB 8; Length 191028;
Best Local Similarity 58.8%; Pred. No. 1,2e-30;
Matches 276; Conservative 0; Mismatches 187; Indels 6; Gaps 1;
```

QY	292	TGAGGAGACACAAATTCGATATTTGTGGAGCGAAGCGAAGTAATAGAGTTG	351
DB	33513	TAAATGAAACATCTTATCTATGCTCTAATCAAGCGAGGGAGCTTCAAGAAATAT	33454
QY	352	CAGGAGTGTTCGAAAGGAGATACAAATATTGAGAGCTTGTGCTGTAACAGAG	411
DB	33453	CCGATTAATTTGCGAGCGAGGCTTAAATAGCAACGCTTGTGAGACACAGACAG	33394
QY	412	ACANG-----CTTATTCACCATAGTGTCTGTGGAAGTGAAGGGTACTTCAGCAGG	465
DB	33393	AAATTTGAGTTTTCAGAAATTCAGATGCTGTTCAAGGAGATTAATAGCAATTCAGCAG	33334
QY	466	TCATCGAGCAACGCCAAGCGTGTATGTTCAAGGTTGAAGATATTCGAAGTGAGC	525
DB	33333	TAACTAACAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	33274
QY	526	CGCAAGTGAAGCGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG	585
DB	33273	CTTCCCGTTGAGAGAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	33214
QY	586	AGATCATGTGGCTAGTTCACACATTCAGACGACGAGTGTGAATATATAGCGGAACATGCAT	645
DB	33213	AAGCTCTGGAATTTGTAAAAATTTTTCAGACCAATGTATGATATATGCGGAAGATTAA	33154
QY	646	TGACTATCGAGTAACGAGATCGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG	705
DB	33153	TAAATTTGTAAGTACGAGACCTTGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	33094
QY	706	AGTTTCAGATCAGAGATTTGTAAAGACAGAAAGATAGCAGTGAAG	754

Db 33093 AATTCGGCATTATTGAATAGTAGAAGCTGGAAAGATTTCGTAGTAAG 33045

RESULT 12
D63676/c
LOCUS
DEFINITION
CYANIDIUM CALDARIUM trnC, trpA, trnT, rps4, trxm, rbcl, rbcS, cfxQ,
trnC, trnL and llyH genes, complete cds.
D63676
D63676.1 GI:2541887
trnA-leu, trnA-cys, trnA-thr, trnA-lys, tryptophan synthase alpha
subunit, thiorodoxin m, ribosomal protein S4, ribulosebiphosphate
carboxylase small subunit, ribulosebiphosphate carboxylase large
subunit, acetylacate synthase, cfxQ protein.
CYANIDIUM CALDARIUM (strain:RK-1) DNA.
SOURCE
ORGANISM
Eukaryota; Rhodophyta; Bangliophyceae; Porphyridiales;
Porphyridiaceae; Cyanidium.
1 (sites)
Ohta,N.
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
Direct Submission
Submitted (31-JUL-1995) Niji Ohta, Waseda university, Research
center for Human Sciences; 2-579-15 Mikajima, Tokorozawa, Saitama
359, Japan (Tel:0429-47-6835, Fax:0429-48-4314)
Location/Qualifiers
1. 7550
/organism="Cyanidium caldarium"
/strain="RK-1"
/db_xref="taxon:2771"
12.. 83
/gene="trnC"
12.. 83
/gene="trnK"
/product="trnK"
108.. 836
/gene="trnA-Lys"
/gene="trpA"
108.. 836
/gene="trpA"
/codon_start=1
/product="tryptophan synthase alpha subunit"
/protein_id="BAA22823.1"
/db_xref="GI:2541888"
/translation="MFTAYLVYAGAPDLNTTKQALNLANDGADVLEIGVPSDPLADG
MILQKASQALNKRFRLEQLMHLAEILPVPVILAYNQIFHYGVEKWTTLVNLK
VKALIVPDLPEEALILRAACARHHLHMTLSPPTKPVRAQOLALACDMLYLSRT
GVTGVDAHFNDQIPNMIIEALKQVTTPIALGFIHQKQLQLVKSAGDVGIIIGTACM
OILLEKGVOLTEWISAMKSSYP"
800.. 2023
/note="ORF407"
/codon_start=1
/protein_id="BAA22824.1"
/db_xref="GI:2541889"
/translation="MDKCDKILIPYGATLAEISKQTNIPITEFIRVYFTQGMITLN
DILDELMLFOAKYGFENKQMSSEKSEEDVDIAIMAVERLDEGVMSQWSTI
STIDRAEVLQMPKALALVHLQOIQDKDILHHRGASIMVVEKRDKLGVVTHLLK
SLIKGDHLEAGVSGKIRALNVKQOQPYTEAKAGIIVHLGTGNOVLATAIINKGAK
EKENTSNQIRIHLCISKGNLIVSESSQLENNVCYLVQEKMLLSAEGSIRMOQIR
LAODTNSILVAMQMTARKOVVEAYDEQILITRLQOIEAOPISITISKAVKOVER
OGTIAGCIYTKGVCFTGMVCQILRGHMLKATAHINIKIHQEVKVKASQEGCISMK
EMNDQVDEIVLIN"
complement(2023.. 2095)
/gene="trnT"
complement(2023.. 2095)
/gene="trnT"
/product="trnA-Thr"
2096.. 2674
/gene="rps4"
/gene="rps4"
2096.. 2674
CDS

/gene="rps4"
/codon_start=1
/product="ribosomal protein S4"
/protein_id="BAA22825.1"
/db_xref="GI:2541890"
/translation="MSRYLGPVRVYIRIRLIGALPALTNKSPPKRTIAPGBHAHRTKLS
EFAVLOAEOKQLQYVYGITNSQLARFQARSKSTGLALTMTLETRLDLVYAGF
APTLPAAROLVNHGHVKNQKKTITASNACENHVEKSSSPKPEPELYIQLSN
GLITVTOPYQKEMLAFFVVELLYEYTRVGA"
complement(2671.. 3432)
/note="ORF253"
/codon_start=1
/protein_id="BAA22826.1"
/db_xref="GI:2541891"
/translation="MDYYIFEFRLSEPOPTVQWSIAMQVSPVQIQHLLSRQVSI
HLPTFKSCGVIDYQPHQFKSNPKIHDAROSILFACPITYVNVNPKOELLSPRK
WSEEQAPTEFNNVYIDCVFROSPTLYLFFHPDQAQNTLOSVLNNPERKINSVOE
GDLFTFLKFNADVAVLVLPHLKQMCQMGVGTVEFVQAKDREMLFLSKONAOQFV
KMKRFLSCQDVLSLELADANQARCVLPLS"
complement(3608.. 3913)
/gene="trxm"
complement(3608.. 3913)
/gene="trxm"
/codon_start=1
/product="thiorodoxin m"
/protein_id="BAA22827.1"
/db_xref="GI:2541892"
/translation="MQIDELSPDTFVLSHOLVLDFFNAPNCGPCRMIAPIIDLEAOE
LDIKVAKMTDNPMLATEYGIKRSIPTLMLEFKGRLDVLIGAVFKSTLMNTEYTL"
4046.. 5512
/gene="rbcl"
4046.. 5512
/gene="rbcl"
/codon_start=1
/product="ribulosebiphosphate carboxylase large subunit"
/protein_id="BAA22828.1"
/db_xref="GI:2541893"
/translation="MAQSVQERTRLNKRRESGVIPYAKMGYMDPNYVYKSDVLALF
RYTPQGVDPTEAANAAGESSATMTVWTDLTACDYYRAKAKVDVADPNEADQYF
VYIAADIDFEEGSIANTPASTIGVYFGKAVKGLRLEPMRIPEYIKTPQGPATGVI
VERERLNKTRGRLDGTCKKGLGSKNKGRRVYBSLKGGLDLDDEINQSPRRW
RDRFLYMEGVNRASASGEVGSYLVNTVAATMEETVYKAEFAKEGVATIMIDLIG
YTAIQSMVAVSRENMMILHLHRAGNSTYSROKHGINFVLSKMMMAVGDHIIAGTV
VGKLEGDPIYIIGFYNTLLPKLEINLPGLGFEDNMWASIRKVMVAASGSHIAGQMQ
LIHYHGEDVLOFGGTTIGHPDGIGAGIAGANVAESMLIANEGRDYFNEGPQLIRD
AKNCGPIQIATLDMKDISFNYSITSDTSFVETPIANV"
5518.. 5934
/gene="rbcS"
5518.. 5934
/gene="rbcS"
/codon_start=1
/product="ribulosebiphosphate carboxylase small subunit"
/protein_id="BAA22829.1"
/db_xref="GI:2541894"
/translation="MKVTQGTFCYLPDLTDEQIRKQLOYAINNGMALSVETVDDPDR
NSYEMWMLPLFDVDRDPATVMEFQVEQRCQRYNNYIKVLAFFNSTGIESTALSTFVNR
PVHEPGFVLRKIESNDRVQRYVHVSATEKPEGSRY"
5991.. 6851
/gene="cfxQ"
5991.. 6851
/gene="cfxQ"
/codon_start=1
/product="cfxQ protein"
/protein_id="BAA22830.1"
/db_xref="GI:2541895"
/translation="MKVNLQTEFFKQTOIQVLDLDDRELIGLQAVKTRIRETALLLV
DRRLKLGSSSPGLHMSFTGSPGKTTVATKMAADILYRGLYIKKGLITVPRDDL
VGQYIGHTAPKTKOYLKNNMGVLEIDEAYLYLRDNEBRYGAEALIELLYQVEMQRD
DLVITFAGYKEMKDRFASNPGLSSRVANHVMPRYTPELIMIGKIMIQEOQYQFTP
EAEVFTQYIQRRMQDPHFANARSVRNALDRARLQANRITFPTPQNLTKDVLVIAQ
EDILKRLFQ"
6862.. 6933
/gene="trnC"
/gene="trnC"

```

tRNA      6862..6933
          /gene="trnc"
          /product="tRNA-Cys"
gene       6938..7023
          /gene="trnl"
          /gene="trnl"
          /gene="trnl"
          /product="tRNA-Leu"
          /complement(7024..7524)
          /gene="llvh"
          /complement(7024..7524)
          /gene="llvh"
          /codon_start=1
          /product="acetylactate synthase"
          /protein_id="BAA22831.1"
          /db_xref="GI:2541896"
          /translation="MKHTLSVLEDEAGVLRIRAGLFARSPNIESLAVGREYQGIS
          RITKVPEDDRIEOLMKLYKPIFIREVNLTPCYERELIMKKVANSDFREIL
          DMANIFRARIIVDIAEDLMLEVTGDPKRVALLQYLAKEGLVEVARTGKIALKRSSKY
          NTEWLR"
CDS
gene
CDS

```

BASE COUNT 2486 a 1367 c 1524 g 2173 t

ORIGIN

Query Match 8.8%; Score 147.2; DB 8; Length 7550;
 Best Local Similarity 56.0%; Pred. No. 1.2e-30;
 Matches 278; Conservative 0; Mismatches 218; Indels 0; Gaps 0;

```

QY 999 CATTACTCTATCATTTGCTTAAATGATATTTCCAGAGGTTCTTAATTTGTGACTGGT 1058
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 7518 CATACCTTTATCTGTATTAATGAAAGTGAAGCAAGAGTCTTAATCTGAATAGCGGTTA 7459
QY 1059 TTCGCTCGAAGGGATACAAATATCCAGAGCTTGCCGTAGACATGCTGAACCAAGGCG 1118
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 7458 TTGGCCGACGATGTTTAACATAGAAAGTTTAGCGGTAGACCCCGCAACAGTGGCG 7399
QY 1119 ATTTACGATTCACACATTTATCTGCAACAGATGAATCGGTACGCAAAATTTGTGCA 1178
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 7398 ATCTCTAGAAATTCATCGTGTGCTGGAGATGACAGAACATAGAACAAATGATGAAA 7339
QY 1179 CAACCTTACAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1238
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 7338 CAATTTACAAATTTCTCCCAATTTTCAAGTAGAAATTTGACACAGTGCCTGTGTA 7279
QY 1239 GAAAGAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1298
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 7278 GAAAGAGATTAATGATTAATGAAAGTGAAGCTAATGACACAAAGAGAAATCTTA 7219
QY 1299 GACATTTGATTAATTTTCAAGGCTTAAGCTGTTGACGATCTGATCACAATTTACTTTG 1358
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 7218 GATATGGCCACATCTTTCGTCTGTATGATGATGATGATGATGATGATGATGATGAT 7159
QY 1359 CAGCTTACTGGGATCTTACAAAGATGTTGACATGCAAGGTTATTTGAGCCCTATGCT 1418
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 7158 GAAGTCACTGGGATGCCGCAAAATGTTGCTTTAGAACAAATTAATTTACTTAATTTGGA 7099
QY 1419 AATATGAGTGTGCAAGAACCGGCTGCTGTGCAATTTGCTGTAATCGGAGTGAAGTCC 1478
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 7098 TTAGTGGATGTGCTTAGAACCGGTTAAATTTGCTTTAAACATCATCAAAAGTAACT 7039
QY 1479 AAGTACCTTGTGAT 1494
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 7038 GAATGTTACGTTAAT 7023

```

RESULT 13
 AE011091/c 11759 bp DNA linear BCT 03-Apr-2002
 LOCUS Methanosarcina acetivorans str. CZA, section 436 of the
 DEFINITION complete genome.
 ACCESSION AE011091 AE010299
 VERSION AE011091.1 GI:19917861
 KEYWORDS Methanosarcina acetivorans CZA.
 SOURCE

ORGANISM Methanosarcina acetivorans CZA
 Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
 Methanosarcinaceae; Methanosarcina.

REFERENCE 1 (bases 1 to 11759)
 AUTHORS FitzHugh, J. E., Nusbaum, C., Roy, A., Endrizzi, M. G., MacDonald, P.,
 Allen, N., Naylor, J., Stange-Thomann, N., Dearliano, K., Johnson, R.,
 Linton, L., McEwan, P., McKernan, K., Talamas, J., Tirrell, A., Ye, W.,
 Zimmer, A., Barber, R. D., Cann, I., Graham, D. E., Grahame, D. A.,
 Guss, A., Hedderich, R., Ingram-Smith, C., Kuetner, C. H.,
 Krzycki, J. A., Leigh, J. A., Li, W., Liu, J., Mukhopadhyay, B.,
 Reeve, J. N., Smith, K., Springer, T. A., Umeyan, L. A., White, O.,
 White, R. H., de Macario, E. C., Ferry, J. G., Jarrell, K. F., Jing, H.,
 Macario, A. J. L., Paulsen, I., Pritchett, M., Sowers, K. R.,
 Swanson, R. V., Zinder, S. H., Zinder, E., Metcalf, W. W. and Birren, B.
 The Genome of *M. acetivorans* Reveals Extensive Metabolic and
 Physiological Diversity
 JOURNAL MEDLINE PUBMED
 GENOME RES. 12 (4), 532-542 (2002)
 11932238
 2 (bases 1 to 11759)
 AUTHORS Birren, B.
 DIRECT SUBMISSION
 TITLE Submitted (20-MAR-2002) Center for Genome Research, Whitehead
 INSTITUTE, Nine Cambridge Center, Cambridge, MA 02141, USA
 FEATURES
 source location/Qualifiers
 1..11759
 /organism="Methanosarcina acetivorans CZA"
 /strain="CZA"
 /db_xref="taxon:188937"
 /complement(1002..1544)
 /gene="MA3789"
 /complement(1002..1544)
 /gene="MA3789"
 /codon_start=1
 /transl_table=11
 /product="nitroreductase"
 /protein_id="AA07140.1"
 /db_xref="GI:19917862"
 /translation="MPVTGKQVDTITLNRSSVREFTRPISKEINTILNAGRRAPS
 GILNQPWFVAVRNPTIOLKSECHYSIGVAGALLIAELDHEHSNRRKDQATG
 ASIONMLACGSLSGAVLGEILNQHEKVNLSIGCSSKLEIMAVLAIGEPAPGRKTS
 TRKALPELVPEEKYGOKWEE"
 /complement(1945..2952)
 /gene="MA3790"
 /note="MA3790"
 /complement(1945..2952)
 /gene="MA3791"
 /codon_start=1
 /transl_table=11
 /product="ketol-acid reductoisomerase"
 /protein_id="AA07141.1"
 /db_xref="GI:19917863"
 /translation="MATIIDNETTFDALDKDTIATIMYGSQGHANFNLHESGLNV
 VGLRKGSSMAKAEESDGLKMTVEEAAKAAVIMILDEKQASVYSOIEPSLEAG
 ALVPHGNHINHOIVPPKDDVDEWAAKGGHIVRYRTGIGVGLIAYQATRG
 ARDLALAAKGIATRAVYETPREPEETDLREGQVNDLGGSLILITAEVLEAG
 YQPEMAVETCHEKLVLDLTTEGGERMHHSVNTAEYGGTVEGPRTINESREMY
 EALARIONGEFAKEFEVLGMVNHPLKMERQEKHPLEVVGKEIRANIPLNKKIDD
 D"
 /complement(3314..3799)
 /gene="MA3791"
 /complement(3314..3799)
 /gene="MA3791"
 /codon_start=1
 /transl_table=11
 /product="acetylactate synthase, small subunit"
 /protein_id="AA07142.1"
 /db_xref="GI:19917864"
 /translation="MKHTLAUVENKSGVLSVRSRGYNIIDSLAVGTEDPDIS
 RITTVHGDHVLQVTKLIDIVKIVSDIGDDAVARELALIKVSADVNTAEI
 OIANIFRARIIVDVAKPSMTVEVTGDESKIQAIERKLRFQGIKEWVRGKIAMVGPVK
 V"

gene
 CDS

SOURCE
ORGANISM
Chloroplast *Guillardia theta*.
Eukaryota; Cryptophyta; Cryptomonadaceae; *Guillardia*.
REFERENCE
AUTHORS
TITLE
1 (bases 47701 to 48415)
Douglas,S.E. and Durnford,D.G.
The small subunit of ribulose-1,5-bisphosphate carboxylase is
plasti-encoded in the chlorophyll c-containing alga *Cryptomonas*
phi
JOURNAL
MEDLINE
93357429
2562756
2 (bases 18535 to 19351)
Douglas,S.E. and Durnford,D.G.
Sequence analysis of the plastid rDNA spacer region of the
chlorophyll c-containing alga *Cryptomonas phi*
DNA Seq. 1 (1), 55-62 (1990)
JOURNAL
MEDLINE
92119320
2132959
3 (bases 43739 to 44938)
Douglas,S.E. and Durnford,D.G.
Nucleotide sequence of the genes for ribosomal protein S4 and
tRNA(Arg) from the chlorophyll c-containing alga *Cryptomonas phi*
Nucleic Acids Res. 18 (7), 1903 (1990)
JOURNAL
MEDLINE
90245587
2336372
4 (bases 34539 to 35380)
Reith,M. and Douglas,S.
Localization of beta-phycoerythrin to the thylakoid lumen of
Cryptomonas phi does not involve a signal peptide
Plant Mol. Biol. 15 (4), 585-592 (1990)
JOURNAL
MEDLINE
91338697
2102376
5 (bases 45872 to 47981)
Douglas,S.E., Durnford,D.G. and Morden,C.W.
Nucleotide sequence of the gene for the large subunit of
ribulose-1,5-bisphosphate carboxylase/oxygenase from the
chlorophyll c-containing alga *Cryptomonas F*: evidence supporting
the polyphyletic origin of plastids
J. Phycol. 26, 500-508 (1990)
JOURNAL
MEDLINE
92099311
1757997
6 (bases 110917 to 113854)
Douglas,S.E.
Unusual organization of a ribosomal protein operon in the plastid
genome of *Cryptomonas phi*: evolutionary considerations
Curr. Genet. 19 (4), 289-294 (1991)
JOURNAL
MEDLINE
9130343
1868578
7 (bases 40675 to 42376)
Douglas,S.E. and Turner,S.
Molecular evidence for the origin of plastids from a
cyanobacterium-like ancestor
J. Mol. Evol. 33 (3), 267-273 (1991)
JOURNAL
MEDLINE
92099311
1757997
8 (bases 96129 to 98906)
Wang,S.L. and Liu,X.Q.
The plastid genome of *Cryptomonas phi* encodes an hsp70-like
protein, a histone-like protein, and an acyl carrier protein
Proc. Natl. Acad. Sci. U.S.A. 88 (23), 10783-10787 (1991)
JOURNAL
MEDLINE
92073372
1961745
9 (bases 106789 to 108216)
Douglas,S.E.
A secY homologue is found in the plastid genome of *Cryptomonas phi*
FEBS Lett. 298 (1), 93-96 (1992)
JOURNAL
MEDLINE
92183838
1544427
10 (bases 42198 to 44153)
Douglas,S.E. and Reith,M.E.
A boch1 homology, encoding a subunit of Mg chelatase, is located on
the plastid genomes of red and cryptomonad algae
J. Mar. Biotechnol. 1, 135-141 (1993)
JOURNAL
MEDLINE
91111111
11 (bases 82327 to 84479)
Douglas,S.E. and Murphy,C.A.

TITLE
JOURNAL
MEDLINE
93357429
2562756
2 (bases 18535 to 19351)
Douglas,S.E. and Durnford,D.G.
Sequence analysis of the plastid rDNA spacer region of the
chlorophyll c-containing alga *Cryptomonas phi*
DNA Seq. 1 (1), 55-62 (1990)
JOURNAL
MEDLINE
92119320
2132959
3 (bases 43739 to 44938)
Douglas,S.E. and Durnford,D.G.
Nucleotide sequence of the genes for ribosomal protein S4 and
tRNA(Arg) from the chlorophyll c-containing alga *Cryptomonas phi*
Nucleic Acids Res. 18 (7), 1903 (1990)
JOURNAL
MEDLINE
90245587
2336372
4 (bases 34539 to 35380)
Reith,M. and Douglas,S.
Localization of beta-phycoerythrin to the thylakoid lumen of
Cryptomonas phi does not involve a signal peptide
Plant Mol. Biol. 15 (4), 585-592 (1990)
JOURNAL
MEDLINE
91338697
2102376
5 (bases 45872 to 47981)
Douglas,S.E., Durnford,D.G. and Morden,C.W.
Nucleotide sequence of the gene for the large subunit of
ribulose-1,5-bisphosphate carboxylase/oxygenase from the
chlorophyll c-containing alga *Cryptomonas F*: evidence supporting
the polyphyletic origin of plastids
J. Phycol. 26, 500-508 (1990)
JOURNAL
MEDLINE
92099311
1757997
6 (bases 110917 to 113854)
Douglas,S.E.
Unusual organization of a ribosomal protein operon in the plastid
genome of *Cryptomonas phi*: evolutionary considerations
Curr. Genet. 19 (4), 289-294 (1991)
JOURNAL
MEDLINE
9130343
1868578
7 (bases 40675 to 42376)
Douglas,S.E. and Turner,S.
Molecular evidence for the origin of plastids from a
cyanobacterium-like ancestor
J. Mol. Evol. 33 (3), 267-273 (1991)
JOURNAL
MEDLINE
92099311
1757997
8 (bases 96129 to 98906)
Wang,S.L. and Liu,X.Q.
The plastid genome of *Cryptomonas phi* encodes an hsp70-like
protein, a histone-like protein, and an acyl carrier protein
Proc. Natl. Acad. Sci. U.S.A. 88 (23), 10783-10787 (1991)
JOURNAL
MEDLINE
92073372
1961745
9 (bases 106789 to 108216)
Douglas,S.E.
A secY homologue is found in the plastid genome of *Cryptomonas phi*
FEBS Lett. 298 (1), 93-96 (1992)
JOURNAL
MEDLINE
92183838
1544427
10 (bases 42198 to 44153)
Douglas,S.E. and Reith,M.E.
A boch1 homology, encoding a subunit of Mg chelatase, is located on
the plastid genomes of red and cryptomonad algae
J. Mar. Biotechnol. 1, 135-141 (1993)
JOURNAL
MEDLINE
91111111
11 (bases 82327 to 84479)
Douglas,S.E. and Murphy,C.A.

COMMENT
STRUCTURAL, transcriptional and phylogenetic analyses of the atpB
gene cluster from the plastid of *Cryptomonas F* (Cryptophyceae)
J. Phycol. 30, 329-340 (1994)
REFERENCE
AUTHORS
TITLE
12 (bases 98901 to 114602)
Wang,S.L., Liu,X.Q. and Douglas,S.E.
The large ribosomal protein gene cluster of a cryptomonad plastid:
gene organization, sequence and evolutionary implications
Biochem. Mol. Biol. Int. 41 (5), 1035-1044 (1997)
JOURNAL
MEDLINE
9137835
92783757
13 (bases 61067 to 68605)
Leitsch,C.E.W., Kowallik,K.V. and Douglas,S.E.
The alpha gene cluster of a cryptomonad, *Guillardia theta*: A piece
in the puzzle of chloroplast genome development
J. Phycol. (1998) in press
JOURNAL
MEDLINE
92119320
2132959
14 (bases 1 to 121524)
Douglas,S.E. and Penny,S.L.
The plastid genome of the cryptophyte alga, *Guillardia theta*:
complete sequence and conserved syntenic groups confirm its common
ancestry with red algae
J. Mol. Evol. 48 (2), 236-244 (1999)
JOURNAL
MEDLINE
99128221
9929392
15 (bases 1 to 121524)
Douglas,S.E.
Submitted (08-JAN-1998) Institute for Marine Biosciences, National
Research Council, 1411 Oxford Street, Halifax, Nova Scotia B3H 3Z1,
Canada
On or before Sep 15, 1998 this sequence version replaced gi:11396,
gi:11297, gi:18103, gi:18281, gi:11383, gi:11407, gi:12539,
gi:136730, gi:11300, gi:398949, gi:2661180.
FEATURES
source
1..121524
/organism="Guillardia theta"
/organelle="plastid:chloroplast"
/db_xref="taxon:55529"
/note="plastid"
/complement(5..86)
/gene="crn1(gua)"
/complement(5..86)
/gene="crn2(gua)"
/product="tRNA-Tyr"
/note="codon recognized: UAC"
/anticodon="(pos:50..52,aa:Tyr)
/complement(181..573)
/gene="rpl19"
/complement(181..573)
/gene="rpl19"
/codon_start=1
/product="ribosomal protein L19"
/protein_id="AAC35594.1"
/db_xref="GI:3602933"
/translation="MEIILHLYEYVNMKSNKNIOLVESNHSNKTLPLOLVGDTVRIG
VSTEGNKRIQATGEVLIISKNSGINTITIRVLAQIGIERVYLLHSPKLSIEIL
RKSIRRAKIVLYLRSGKATRRLOQRN"
729..3188
/gene="clpC"
/gene="clpC"
729..3188
/codon_start=1
/product="clp protease ATP binding subunit"
/protein_id="AAC35595.1"
/db_xref="GI:3602934"
/translation="WEFRTKAIKIVIMLAQEARRLGHNFVTEQILLIGEGTGI
AAKVLKSGVNLKDAVEREVIIGRSGFVAVEIPFPRAKRVLELSLEEROLGHNH
IGTEHLLGLIREGGEVAVRLEMLALDTKRVQVIRLGDPAVSATNQTGKTP
TLERGSNLTOKAAGKIDPVIYGRKTEIRVLIQIGRTKNPILIGEGVGTAAIE
GLAQRINRDPDITIEDKRVYTLIDGLIVAGTRYGEEERLKTITIDIRANVIV
IDEVTLTGAGAGSADIDANILKPAKRGEOCTGATITLLEIRHLEKDSLEERFO
PVWGEPSVEETITELLYGIRDORYEKHNHLVLSDEALSAARFADYIADRPLPKAID
LIDEGSRVRLMNSQLPPAARLEDELEIKKQDEAVASODFEYAGOLDRREMEIKA
OIAAIAHSKKKGDEENTKEVSVTEDIDIAQIVAAWTGIVPNKMTRESEKLLQMEETLH

```
GRIGODEAVAVASKAIRARVGLKNPNRIASFISGPTGVGKTELLTKALASYEFGS
FEAWRLDMSEYMERHTVSKLIGSPGYVNGGLTSEVRRPTVVLFPBIEGKH
PEVFNLLDILEDRITDLSKGRVDEKNTLLITSVSGKVIKGGGGLPDLSEDT
FEOYGRKALVNEBELKQVREPFLNLDITIVYRQLTKEVCEIAFIMAKETPRISE
KGIQLEVTARFHTHLINECYNPICYARPIRRAYMRLEDTLSEBFLAEIKKGDIAVY
DVDDDKRVAVLIGEKLELVAN"
/ gene="trnf(gaa)"
/ 3316 . .3388
/ gene="trnf(gaa)"
/ 3316 . .3388
/ product="trn(gaa)"
/ note="codon recognized: UUC"
/ anticodon=(pos:3349. .3351,aa:phe)
/ complement(3402. .5201)
/ gene="dnab"
/ complement(3402. .5201)
CDS
/ gene="dnab"
/ codon_start=1
/ transi_except=(pos:complement(5199. .5201),aa:met)
/ product="replication helicase subunit"
/ protein_id="AAC3596.1"
/ db_xref="GI:3602935"
/ translation="MKVTKFYNLKAKEHLIIQILVEKDFEIQITKIDTQVYFQSHQ
ILYRSINILHTKRVLINLPNLISLHSELVSETVINYLITLFRNDRFRLINISLK
VLIDNFIRELQNSCAKIVSLTFENFLSVETLQKSNLISINTYKNSLKSLSQ
LLETLIEDKTNRSTHVLTFEFDLHILVGLQKSLIITAGRSPMGKAFMLSVRN
VDIOSEFIVIFSLEMSKQILYRLISNENATSRREGNISISEMELNMTLVS
NINIVLDENNDIVLDIOLSKLSLOOIOYDIGLIVDIQILOQYKRSQRHIELSYI
TRKLITAKDFNPLVLYLSQLSKNVEERLNKRPILSLDESGLSTATQPFYLSDN
VNSKLTLSKVSNNHLYLSATLELLIPEQYNRIYPIVSLKRELQYGVVYLFDEYS
VIVSTENHYVLLINQMKRILELVDLDVATLDIOYLINNTEDLFSVIVSSVNL
ICMNRINVDPMIPKTNFNFVALLVHNSIEDADAVLELYRDSYVQHNLKQSNIDMC
EVIYAKRHGTCIGMVLNLPNTVSEFMNLIKES"
misc_feature
3594. .4073
/ note="signal element intein; encodes putative intein
spliced out of the replication helicase subunit"
gene
5361. .5433
/ gene="trnW(cca)"
5361. .5433
/ gene="trnW(cca)"
/ product="trnW(cca)"
/ note="codon recognized: UGC"
/ anticodon=(pos:5394. .5396,aa:Trp)
5588. .6013
/ gene="rp111"
5588. .6013
/ gene="rp111"
/ gene="rp111"
/ codon_start=1
CDS
```

Query Match 8.7%; Score 145; DB 8; Length 121524;

Best Local Similarity 56.0%; Pred. No. 7.1e-30;

Matches 274; Conservative 0; Mismatches 215; Indels 0; Gaps 0;

```
QY 999 CATACCTCATCTGCTTGTAAATGATATTCACAGAGCTTCTAATATTGACTGCTGTT 1058
||||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 50000 CATACATTAATACAGTTTGTGTAAGATGAACTGCGTGTAACTCGTATTGCTGTTTA 49941
QY 1059 TTGGCTCGAAGGGATACATATCCAGAGCTTGCGCTAGACATGCTGAACCAAGGCC 1118
||||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 49940 TTGGCTGCTGCTGTTCAATATCGAAAGTTTAGCACTGTGACCAATGAAAAATTAGA 49881
QY 1119 ATTTACGCACTTACACAGATTAATACCTGCAACAGATGAATCGGTCAGCAAAATGTCGAC 1178
||||| ||||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 49880 ATTTACAGTATTACTATAGTGTGTTCCATCGGATGAAGACATATTAGCAATTAACAAA 49821
QY 1179 CAACCTTACAACCTGAGATGTGCATGAGGTCACATGATCTTACTCATTTTGCATTTTCT 1238
||||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 49820 CAACCTTATAGCTTGTATTATATTAAAGTTGAAGATTAATACGAATCTACACGTT 49761
QY 1239 GAAAGGAATGATGCTGATTAAGATTTGCCGTGACCGCTGCTGAGAAAGATGTCCTG 1298
|| | ||||| || | ||||| | | | | | | | | | | | | | | | | | | | | | | |
Db 49760 GAGCGAAGACTGTTGATTAAATTAACGAGTTTCAACATTTAGAGAGACTGAAGCATTA 49701
```

```
QY 1299 GACATTGCTAGTATTTTTCAGGCGCTAAAGCTGTTGACATATCTGATCATCACAAATTACTTTG 1358
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 49700 GATATAATTATATATTTCGTGCTAAGGTTGTAGACCTATCTAGAGATTTTTTAATAATT 149641
QY 1359 CAGCTTACTGGGAGCTATGACAGAGATGTTGCACCTGCCAAGGTTATTGACGCCCTATGCT 1418
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 49640 GAAGTTACAGCTGATCCTGTTAAATTTGTTCCATAGAGCAAAATTTTGTCAAAATTTACT 49581
QY 1419 ATATGAGAGTTGCAAGAACCGGCTGCTGCGCATTTGGCTGCGTAATCGGAGGAGCTCC 1478
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 49580 ATATCTGAGATGACTAAGAGACTGAAAAATTTCTTGTGTACGCAATCTAATGTTAATACC 49521
QY 1479 AACTACTT 1487
| | |
Db 49520 CGTTCTCTT 49512

RESULT 15
AP003597 342850 bp DNA linear BCT 28-NOV-2001
LOCUS Nostoc sp. PCC 7120 DNA, complete genome, section 17/19.
DEFINITION AP003597 BA000019
ACCESSION AP003597 GI:17133729
VERSION AP003597.1
KEYWORDS Nostoc sp. PCC 7120 DNA.
SOURCE Nostoc sp. PCC 7120
ORGANISM Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
REFERENCE 1
AUTHORS Kaneo,T., Nakamura,Y., Wolk,C.P., Kuritz,T., Sasamoto,S.,
Watanabe,A., Iriyuchl,M., Ishikawa,A., Kawashima,K., Kimura,T.,
Kishida,Y., Kohara,M., Matsumoto,M., Matsuno,A., Muraki,A.,
Nakazaki,N., Shimo,S., Sugimoto,M., Takezawa,M., Yamada,M.,
Yasuda,M. and Tabata,S.
Complete genomic sequence of the filamentous nitrogen-fixing
cyanobacterium Anabaena sp. strain PCC 7120
DNA Res. 8 (5), 205-213 (2001)
2 (bases 1 to 342850)
Kaneo,T.
Direct Submission
Submitted (02-MAY-2001) Takakazu Kaneo, Kazusa DNA Research
Institute, The First Laboratory for Plant Gene Research; Yana
1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail:kaneoko@kazusa.or.jp.
URL:http://www.kazusa.or.jp/cyanobase/,
Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)
Location/Qualifiers
1. 342850
/organism="Nostoc sp. PCC 7120"
/db_xref="taxon:103690"
/ note="synonym:Anabaena sp. PCC7120"
complement(174. .1478)
/ gene="a114593"
/ complement(174. .1478)
/ gene="a114593"
/ complement(174. .1478)
/ note="ORF ID:a114593
hypothetical protein"
/ codon_start=1
/ transi_cable=11
/ protein_id="BAB76292.1"
/ db_xref="GI:17133730"
/ translation="MDKSIQLVDELPSNITVVKVNLADYVAPGOMONLVGFNDTIR
AVTGFDARVQIKIRDRATSIYNDPOGQSVKIXQRTDKADTAMATRALANKVEK
IGFLSFLSNITPRKAVYOSIDLVSKAIEIIFYECLNIGTPONRPNRANSLANNQDA
SLIRNVALVCLDGIIPLGDFLSKIGTQISGAESENSTIONPVLVNSLPGSNFTDK
IGFINQGFNAVOGMNNNLVTKTGPQISISHLGNFIADNDLDFVNSAFDQNTNF
EHTGQISVARRLILKAHTLVKEIRIQEFTQPIODSSAVKSDTSQYAAASKIEVDEE
BEDWQIATIEKVDQDFINVIYIGGSSHNEMWGDIDRATDYASSDANGAVGVKKVVC
WSDEETWTATIEKIQGNQYFIHTGTGDSYDENVNDDIS"
gene
1680. .1880
/ gene="asr4594"
1680. .1880
/ gene="asr4594"
CDS
```

```
gene
/feature="ORF_ID:asr4594
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB76293.1"
/db_xref="GI:17133731"
/translation="MIAIAPOEKSPPROVLLYLGEDNFFVEVPSLPGLPGCISOGKTRREA
LTNIEVLQDKGETLPEYOFINI"
1877..2770
/gene="a1r4595"
1877..2770
/gene="a1r4595"
/feature="ORF_ID:a1r4595
unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB76294.1"
/db_xref="GI:17133732"
/translation="MSKKLVDRLLFCFTFVLGFTTITHVQAESAPEYWEFINVDA
VQPGDMIVETQKYNFIRDYNHLRSRYIPLDNVDRIEVSSENGQLLPSITSTG
NHQLLSWEHQLSPESHTEFLKRVIGGLVNNYAOVWEALFSDRKADIKAOVR
VEPEELLNKKIKIFOSGVVADIRVDKTEAVTKOIEPRTGKIQVIEDRAGTKI
ETSGMOSQOPFERLWTLGLVFGIPLLLIYSRNGTGTSSSIIGDSSIDYIGGGG
DYGGGDYGGDYGGGGGGGD"
3082..4473
/gene="a1r4596"
3082..4473
/gene="a1r4596"
/feature="ORF_ID:a1r4596"
/codon_start=1
/transl_table=11
/product="malic enzyme"
/protein_id="BAB76295.1"
/db_xref="GI:17133733"
/translation="MADLTPESSFSVTIRLQIPNRVGMLASITQAIASSGNGLOIDL
IEQRVSLRDLTVDASTEHEFTVQAKAIPDKLVSYDRFTNLRGKISITSR
IPLKVSVDLAMYTPGVGRCVTAIAODSOVYNTIKONTVAIVTDSAVLGNLGP
HAALPYMGKAMLFKEFAGLDAPICLADOTTEIRIQAKNIAPVGGVNLEDIAAPR
CFEIEKRLRDLDPVFDHDOHTAIVTALAFNSLKLVOKSLADIRIVINAGAGV
AIALRLKAGAKETIMCDSKGIISTRTDLNEEKQEPFAVKAOGTAGVQADYEIVG
SAPGVLPPEWYKSAKDAIVFAMNPPIEIOPELISKDVAVIATGRSDYPMQINNVLA
EPGVFRGALDCAQOTITTTMYLEASALIASLVNPSDSLREHIIPSEDERVVTAVAAA
VQRAAREGIIARG"
5270..6853
/gene="a1r4597"
5270..6853
/gene="a1r4597"
/feature="ORF_ID:a1r4597"
/codon_start=1
/transl_table=11
/product="type I restriction-modification system DNA
methylase"
/protein_id="BAB76296.1"
/db_xref="GI:17133734"
/translation="MTEKAARKNNSESEFQRLMKRAADKLRRKNIDAEXKHIVLGIIFL
KYIDAFELHOKLAGEAGAGANPEDRDEYSANVFVPEARMSYLOQAOAPDI
KGTVDLAEALIEQENAKRLGILPKVYGOOKLDOKSLGGLIDLGISITLGAEOAOPD
VLGVVYELGDFALAECKKGGOFTPEISIVKILVEMLEPYNGRVPDPDCCSGGMAVO
SEKPYKNQGRDLDISITGQESNETTYKLCRMNLAKRIGDSNITKMNPEGFLDAHK
DLKADFYIANPFNDSDWGGELLRNDGKMDKDLVPPVGNANFAMVSHFTIHLAPTGS
AGFVLSNGSLSSNTSGEGDIRKALVQKDLVDCIYVLPOLFYNTGIPACLWFLPRYKN
GKNRDRHGEVLFIDASELGYMNVNSSRAFTAEATISKIADTYHEWKKSGGSYRDIKGF
CKSASIAIEIKHNFLVLPGRYVGIPIDEVEDGVSPFEKKSSELTLMALGEGMRGQLLDEE
IKKOLLKYGFTVSELVDVF"
6858..7121
/gene="asr4598"
6858..7121
/gene="asr4598"
/feature="ORF_ID:asr4598
unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB76297.1"
```

```
gene
/db_xref="GI:17133735"
/translation="MMNTQVTLKKLNPVIEKLLKHLAQSORTLEETITSLLEDVTEN
TPITTSKSDWSPGPFQFCAGMGCELLVREPQPEADREPL"
7226..7804
/gene="a1r4599"
7226..7804
/gene="a1r4599"
/feature="ORF_ID:a1r4599
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB76298.1"
/db_xref="GI:17133736"
/translation="MIAVKGKPKLTPEEYFIMEKQOLRHEYLNGEVYAMTGGTONH
GRASNIIFIKVHGLRGCCGVNSDRINIEFKDYVYDPVSVTCDEDRATAIOAO
VPCILVETSPSTASYDRGDKFRLYRRNPSLODVLVDAEKIALIDLKRNKRWMLF
NYQLGDNKIOSIDLSPEFIESVEDIVFEESV"
7855..8145
/gene="asr4600"
7855..8145
/gene="asr4600"
/feature="ORF_ID:asr4600
unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB76299.1"
/db_xref="GI:17133737"
/translation="MIREKNMTQVILLKQNLNPLYLEKLRQAQSHQRTLEETISILE
DVAKEEPPKSSSEGFMDWTLOFRERMOENITFDADPADIRDSVGHDFE"
8211..8573
/gene="a1r4601"
8211..8573
/gene="a1r4601"
/feature="ORF_ID:a1r4601
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB76300.1"
/db_xref="GI:17133738"
/translation="MVKKLTHQLETAGSVYVHMLLFCGLRIVESQKRRLLEYINO
IPLKMTILNDLKAQWMAQERARLSKMGKTPATIDQIASIAYSNLLITVNVSDF
ESFNDLAVENFVNSGEG"
8577..9749
/gene="a1r4602"
8577..9749
/gene="a1r4602"
/feature="ORF_ID:a1r4602"
/codon_start=1
/transl_table=11
/product="type I site-specific deoxyribonuclease chain S"
/protein_id="BAB76301.1"
/db_xref="GI:17133739"
/translation="MSEWKEETLGEIADIIMGQSPGTGTCNNNGQGLPLNGPEFGD
RNPLPTQFTIDPKRIAEGDGLFCVRGSGTGRMMWADQKVAIGGIASIRAKDGLFQ
PYIRAILIEKELSLAVATGSTFPNISDHLNLNIVQIPSKIRIYISNLARIIDEXI
YNIISQNETLEAIQOTLEKHFIDPEEPNADGKRYKSSGAMVSAALCYIPEASVKG
LGOVLTIRKHGVAFGEVYTTETVTKILLTPVNFYIGGGFNSSKYKYSADQSNSEYVL
RRKLATMPLDMSKESGSLGCPATIPDQKGVPLIANDRIGKVEYNNNDKTLTYELLC
REYSHILGSSGSTVHTSPSRICETSFVLPDELIDKFSALATATIDKILISTL"
9944..10465
/gene="a1r4603"
9944..10465
/gene="a1r4603"
/feature="ORF_ID:a1r4603
unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB76302.1"
/db_xref="GI:17133740"
/translation="MIGKLRRLNDRNSFSHIALAESQADQKYNILISKYIDLFEV
KKLSISLIQYKNTLSNTTDLRFLKFDGSHSKKRNHDLVDNNFIRTNIDVNEELRF
CKFIANNQIICLSPAYGYLHNGYPHILEYKQAEPPNYFIEVIAADQPREIKIERNI
FDVSIQIIESLLL"
```


GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 9, 2003, 18:38:17 ; Search time 417 seconds
(without alignments)
9034.996 Million cell updates/sec

Title: US-09-997-900-1
Perfect score: 1673
Sequence: 1 gctcttcagtagcaaaaa.....tgatttggcttcattcg 1673

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N.Geneseq_101002.*
1: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1673	100.0	1673	21	AAA27424
2	875.6	52.3	960	24	ABN98543
3	701.8	41.9	1721	24	ABA98820
4	665.8	39.8	1861	19	AAV55857
5	593.8	35.5	1435	24	ABA98819
6	590	35.3	1813	24	ABA98816
7	537.6	32.1	1287	24	ABA98814
8	419	25.0	4895	21	AAA27425
9	243.4	14.5	611	19	AAV55854

10	201.8	12.1	724	24	ABA98822
11	197	11.8	470	24	ABA98813
12	169.4	10.1	515	24	ABA98817
13	160.8	9.6	484	24	ABA98818
14	149.4	8.9	417	24	ABA98821
15	147.6	8.8	420	21	AAC37251
16	112	6.7	565	24	ABD27618
17	97.8	5.8	547	24	ABA98815
18	92	5.5	230	19	AAZ96290
19	91	5.4	525	22	AA500993
20	91	5.4	3291	22	AA502313
21	91	5.4	10426	22	AA502329
22	90.4	5.4	6211	19	AAV52141
23	90	5.4	501	23	AA555637
24	90	5.4	501	23	AA555660
25	79.8	4.8	13732	21	AAA81684
26	79.8	4.8	34980	21	AAE21611
27	76	4.5	12720	15	AA064211
28	74.4	4.4	516	22	AAH68423
29	74.4	4.4	639	22	AAE71477
30	74.4	4.4	2427	22	AAE54905
31	74.4	4.4	34980	22	AAH68528
32	71	4.2	489	24	ABQ90233
33	64.8	3.9	640681	24	ABA92787
34	64.6	3.9	522	23	AA556283
35	62.6	3.7	489	23	AA551405
36	62.6	3.7	522	23	AA554008
37	62.6	3.7	96109	22	AAE28548
38	62.6	3.7	319630	24	ABQ67194
39	62.6	3.7	3011208	24	ABQ69245
40	61.6	3.7	2286	24	ABQ69168
41	61.6	3.7	6506	24	ABQ71032
42	60	3.6	561	17	AAT45894
43	57.8	3.5	2944528	24	ABA03041
44	56.8	3.4	492	23	AA554309
45	55.4	3.3	492	22	AAH78452

ALIGNMENTS

RESULT 1	AAA27424	standard; cDNA; 1673 bp.
ID	AAA27424:	
XX	11-AUG-2000	(first entry)
XX	Arabidopsis	acetylhydroxy-acid synthase (AHAS) small subunit cDNA.
DE	Herbicide-resistance; acetylhydroxy-acid synthase; AHAS;	
KW	acetolactate synthase; imidazole; sulfonilurea;	
KW	triazolopyrimidine sulfonamide; sulfamoylurea; enzyme;	
KW	pyrimidyl-oxy-benzoic acid; sulfonilcarboxamide; transgenic plant;	
KW	branched-chain amino acid synthesis; ss.	
XX	Arabidopsis sp.	
OS	Key	Location/Qualifiers
XX	Key	42..1517
FT	CDS	/*tag= a
FT		/product= "Arabidopsis AHAS small subunit"
FT		/EC_number= 4.1.3.18
XX	PN	W0200026390-A2.
XX	PD	11-MAY-2000.
XX	PF	28-OCT-1999; 99MO-US25452.
XX	PR	29-OCT-1998; 98US-0106239.

Wheat ALS small su
Corn ALS small sub
Rice ALS small sub
Rice ALS small sub
Wheat ALS small su
Arabidopsis thaliana
Synchocystis sp.
Corn ALS small sub
S. pneumoniae deri
B. subtilis gene e
B. subtilis livBNC
Plasmid PAN267 for
Streptococcus pneu
Streptococcus pneu
Streptococcus pneu
N. meningitidis pa
Neisseria meningit
L. lactis branched
C glutamicum codin
Corynebacterium q1
Nucleotide sequenc
C glutamicum codin
M. capsulatus gene
Buchnera sp. genom
Salmonella typhi D
Klebsiella pneumonia
Klebsiella pneumonia
Genomic fragment #
Listeria innocua C
Listeria innocua D
Listeria monocytog
Listeria monocytog
R. capsulatus acet
Listeria monocytog
Pseudomonas aerugi
DNA of small subun

PA (AMCY) AMERICAN CYANAMID CO.
XX
PI Kakefuda G, Costello C, Sun M, Hu W;
XX WPI: 2000-365633/31.
DR P-PSDB; AAI96203.
XX
PT New polynucleotide encoding eukaryotic acetoaldehyde synthetase
PT small subunit protein for producing transgenic herbicide resistant
PT plants and identifying mutations affecting enzymatic activity of the
PT synthetase
XX
PS Claim 2; Page 47-50; 57pp; English.
XX
XX The present sequence is the cDNA for Arabidopsis acetoaldehyde-acid
CC synthase (AHS) small subunit protein. AHS is also known as
CC acetylaldehyde synthase. This enzyme is needed for branched-chain amino
CC acid synthesis and so is essential for life. Inhibition of this enzyme
CC would lead to plant death and therefore inhibitors would be potential
CC herbicides. Certain herbicides are known to inhibit AHSs:
CC imidazolinones, sulfonylureas, triazopyrimidine sulfonylureas,
CC pyrimidyl-oxo-benzoylureas, sulfamoylureas and sulfonylcarbamides.
CC The present sequence may be used to identify mutant AHSs which are
CC resistant to these herbicides and may be used to create herbicide
CC resistant transgenic plants e.g. dicot and monocot crop plants.
XX
SQ Sequence 1673 BP; 466 A; 328 C; 418 G; 461 T; 0 other;

Query Match 100.0%; Score 1673; DB 21; Length 1673;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1673; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTTCAGTACGCAAAACCTTCGCTTCGTCATGATGCGCCATTTCGTAA 60
DB 1 GTCCTTCAGTACGCAAAACCTTCGCTTCGTCATGATGCGCCATTTCGTAA 60
QY 61 GTCTTCACCATCTATTCGCTTCGTCATGATGCGCCATTTCGTAA 120
DB 61 GTCTTCACCATCTATTCGCTTCGTCATGATGCGCCATTTCGTAA 120
QY 121 TATCCTTCAGTACGCAAAACCTTCGCTTCGTCATGATGCGCCATTTCGTAA 180
DB 121 TATCCTTCAGTACGCAAAACCTTCGCTTCGTCATGATGCGCCATTTCGTAA 180
QY 181 ACCGTGCGATGAAATGGTGAAGATGGAAGATTCGTTAGAAAGCGTGTAGAA 240
DB 181 ACCGTGCGATGAAATGGTGAAGATGGAAGATTCGTTAGAAAGCGTGTAGAA 240
QY 241 TCTCTGATGCGCTTCCTCGAAAGCTTCATCTCGACTCCAAATCGAAGTGAAGC 300
DB 241 TCTCTGATGCGCTTCCTCGAAAGCTTCATCTCGACTCCAAATCGAAGTGAAGC 300
QY 301 ACACAAATTCAGTATTCGTTGAGAGCAAGCGAATGATTAGATTCGAGAGTGT 360
DB 301 ACACAAATTCAGTATTCGTTGAGAGCAAGCGAATGATTAGATTCGAGAGTGT 360
QY 361 TTGCAAGAGAGATACAAATATTGAGAGCTTCGTTGCTGAACAGAGCAAGCTTC 420
DB 361 TTGCAAGAGAGATACAAATATTGAGAGCTTCGTTGCTGAACAGAGCAAGCTTC 420
QY 421 TATTCACCAATAGTTCGTTGAGAGCAAGCTTCGTTGCTGAACAGAGCAAGCTTC 480
DB 421 TATTCACCAATAGTTCGTTGAGAGCAAGCTTCGTTGCTGAACAGAGCAAGCTTC 480
QY 481 AGAAGCTCGTAAATGTTGAGAGCAAGCTTCGTTGCTGAACAGAGCAAGCTTC 540
DB 481 AGAAGCTCGTAAATGTTGAGAGCAAGCTTCGTTGCTGAACAGAGCAAGCTTC 540
QY 541 AGCTGATGCTTTGAAAGTGAATGCAATCCAGATCCAGAGATCATGAGCTAG 600
DB 541 AGCTGATGCTTTGAAAGTGAATGCAATCCAGATCCAGAGATCATGAGCTAG 600
QY 601 TTGACACATTCAGAGCAAGATTTGATGATTCGGAACATGATTCATTCAGAGTAA 660

DB 601 TTGACACATTCAGAGCAAGATTTGATGATTCGGAACATGATTCATTCAGAGTAA 660
QY 661 CTGAGATGCTCGAAGAAATGATGCTGTAAGAAATTTGAAAGTTTCATCATCAG 720
DB 661 CTGAGATGCTCGAAGAAATGATGCTGTAAGAAATTTGAAAGTTTCATCATCAG 720
QY 721 AGATTGTAAGAGCAAGAAATGATGCTGTAAGAAATTTGAAAGTTTCATCATCAG 780
DB 721 AGATTGTAAGAGCAAGAAATGATGCTGTAAGAAATTTGAAAGTTTCATCATCAG 780
QY 781 TTTGGCGATTTTCAGCAGCATTCCTATTCAGATCTCAAGGCAAGCGCTGTTAGTTC 840
DB 781 TTTGGCGATTTTCAGCAGCATTCCTATTCAGATCTCAAGGCAAGCGCTGTTAGTTC 840
QY 841 TTGAGTACCAAAAAAGGCGATTCCTCAAGGCAAGCGCTGTTAGTTC 900
DB 841 TTGAGTACCAAAAAAGGCGATTCCTCAAGGCAAGCGCTGTTAGTTC 900
QY 901 TTTATCCGTTGAGCCATTTTGGACCCAGGTATCTATTCGACGCTCAGTGGG 960
DB 901 TTTATCCGTTGAGCCATTTTGGACCCAGGTATCTATTCGACGCTCAGTGGG 960
QY 961 GACTTCTACTGACGAAGATAGCATGAGTACGCTGATCTATTCATTCCTGTA 1020
DB 961 GACTTCTACTGACGAAGATAGCATGAGTACGCTGATCTATTCATTCCTGTA 1020
QY 1021 ATGATATTCAGAGATTCCTAATATGTCGCTGCTGCTGCTGCAAGGGATACAA 1080
DB 1021 ATGATATTCAGAGATTCCTAATATGTCGCTGCTGCTGCTGCAAGGGATACAA 1080
QY 1081 TCCAGAGCTTGGCCGTTAGAGCATGCTGAACCAAGGCAATTCACCATTCACAG 1140
DB 1081 TCCAGAGCTTGGCCGTTAGAGCATGCTGAACCAAGGCAATTCACCATTCACAG 1140
QY 1141 TACCTGCAACAGATGAAATGCGTCAAGAAATTTGTCAGCAACTTTCACAACTG 1200
DB 1141 TACCTGCAACAGATGAAATGCGTCAAGAAATTTGTCAGCAACTTTCACAACTG 1200
QY 1201 TGCATGAGTCCATGATCTTACATTCATTCGATTTCTGAAAGAAAGATGATCT 1260
DB 1201 TGCATGAGTCCATGATCTTACATTCATTCGATTTCTGAAAGAAAGATGATCT 1260
QY 1261 AGATTGCGTGAAGCGTGCCTGTAAGAAAGATGCTGCAATTCGATTTTCAGGG 1320
DB 1261 AGATTGCGTGAAGCGTGCCTGTAAGAAAGATGCTGCAATTCGATTTTCAGGG 1320
QY 1321 CTAAAGCTGTTGACGATCTGATCACAACAATTTCTTGGACGCTTACCTGGGATCT 1380
DB 1321 CTAAAGCTGTTGACGATCTGATCACAACAATTTCTTGGACGCTTACCTGGGATCT 1380
QY 1381 AGATTGCTGACGCAAGAGTTTATGAGGCGCTATGATATGATGAGTTGCAAGAAC 1440
DB 1381 AGATTGCTGACGCAAGAGTTTATGAGGCGCTATGATATGATGAGTTGCAAGAAC 1440
QY 1441 GTCGTTGCGATTTGCTGTAATTCGGAATGAGTGCATCAAGTCTGCTGATTCCT 1500
DB 1441 GTCGTTGCGATTTGCTGTAATTCGGAATGAGTGCATCAAGTCTGCTGATTCCT 1500
QY 1501 TTTCTTTTAAAGGCTTAAAGCTTGAAGAGTGCATTCATGCAATCAGAAACTTTG 1560
DB 1501 TTTCTTTTAAAGGCTTAAAGCTTGAAGAGTGCATTCATGCAATCAGAAACTTTG 1560
QY 1561 GTAAAGTTTCATTCACAGCTATGACCTCAAGACCAAGACAGAGAGAGTGTGATA 1620
DB 1561 GTAAAGTTTCATTCACAGCTATGACCTCAAGACCAAGACAGAGAGAGTGTGATA 1620
QY 1621 TATGTTTGTGACTTTGTTATGAACAATTTAGCTGATTTTGGGCTTCATTTG 1673
DB 1621 TATGTTTGTGACTTTGTTATGAACAATTTAGCTGATTTTGGGCTTCATTTG 1673

RESULT 2

ABN98543/c
ID ABN98543 standard; DNA; 960 BP.
XX
AC ABN98543;
XX
DT 01-ANG-2002 (first entry)
DE Arabidopsis thaliana expressed polynucleotide SEQ ID NO 311.
XX
XX Arabidopsis thaliana; plant; insecticide; fungicide; transgenic; stress;
KM disease crop; thale cress; tolerance factor; insect; pathogen;
KM nutrition; ds.
XX
OS Arabidopsis thaliana.
XX
PN US200203281-A1.
XX
PD 21-FEB-2002.
XX
PF 26-JAN-2001; 2001US-0770445.
XX
PR 27-JAN-2000; 2000US-178472P.
XX
PA (GORL/) GORLACH J.
PA (ANY/) AN Y.
PA (HAM/) HAMILTON C M.
PA (PRIC/) PRICE J L.
PA (RAIN/) RAINES T M.
PA (YUY/) YU Y.
PA (RAME/) RAMEKA J G.
PA (PAGE/) PAGE A.
PA (MATH/) MATHEW A V.
PA (LEDF/) LEDFORD B L.
PA (WOES/) WOESSNER J P.
PA (HAAS/) HAAS W D.
PA (GARC/) GARCIA C A.
PA (KRIC/) KRICKER M.
PA (SLAT/) SLATER T.
PA (DAVI/) DAVIS K R.
PA (ALLE/) ALLEN K.
PA (HOFF/) HOFFMAN N.
PA (HURB/) HURBAN P.
XX
PI Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;
PI Rameka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;
PI Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;
PI Hurban P;
DR WPI; 2002-400781/43.
XX
XX New Arabidopsis thaliana nucleic acid for identifying homologous genes,
PT producing compositions that modulate the expression or function of its
PT encoded protein, and mapping functional regions of protein
XX
XX Claim 1; SEQ ID NO 311; 49pp + Sequence Listing; English.
XX
XX The invention relates to an Arabidopsis thaliana nucleic acid (I)
CC comprising a sequence capable of hybridising under stringent conditions
CC to a sequence selected from any one of 999 sequences (ABN98233-ABN98231),
CC given in the specification or its fragment. A polypeptide (II) encoded by
CC (I), a transgenic plant (III) comprising an exogenous nucleic acid or a
CC genetically modified cell (IV) comprising an exogenous nucleic acid, is
CC useful for screening a candidate agent for its biological effect. (I) is
CC useful in identifying homologous or related genes, in producing
CC compositions that modulate the expression or function of its encoded
CC protein, mapping functional regions of the protein and in studying
CC associated physiological pathways. (I) is also useful for the genetic
CC manipulation of cells, particularly plant cells. (I) is also useful in
CC screening assays of various plant strains to determine the strains that
CC are best capable of withstanding a particular disease or environmental
CC stress. (II) and (III) are useful for screening of biologically active
CC agents, e.g. fungicides, insecticides, etc., for elucidating biochemical
CC pathways. The screened agents are useful in improved methods of treating

CC crops to prevent or treat disease. (II) are also useful in screening
CC programs to identify agents that mimic or enhance the action of tolerance
CC factors. Such agents are useful in improved methods of treating crops to
CC enhance their tolerance to environmental stress. (I) is also useful
CC for enhancing or inhibiting production of a biosynthetic product in a
CC plant. (III) is useful for identifying other mediators that may induce
CC expression of proteins of interest, for establishing the extent to which
CC any specific insect and/or pathogen is responsible for damage to a
CC particular plant, for identifying other mediators that enhance or induce
CC tolerance to environmental stress, for identifying factors involved in
CC biosynthetic pathways of nutritional, commercial, or medicinal value and
CC for identifying productions of nutritional, commercial or medicinal
CC value. (IV) is useful in the study of genetic functional and regulation,
CC for alteration of the cellular metabolism and for screening compounds
CC that may affect the biological function of the gene or gene products.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?DocId=99909770445.
XX
SQ Sequence 960 BP; 279 A; 221 C; 191 G; 269 T; 0 other;

Query Match 52.3%; Score 875.6; DB 24; Length 960;

Best Local Similarity 98.7%; Pred. No. 1.9e-261;
Matches 894; Conservative 0; Mismatches 9; Indels 3; Gaps 1;

QY 768 GCAACGCTCCATTTTGGCGATTTTTCAGCAGCATCCATCCAGATCCAAAGAGCAAGCG 827
DB 960 GCAACGCTCCATTTTGGCGATTTTTCAGCAGCATCCATCCAGATCCAAAGAGCAAGCG 901
QY 828 CCGTGTAGTGTCTCGAAGATAGCAAAAAGAGCCATGTCCCTCAAAAAGAAATCA 887
DB 900 CCGTGTAGTGTCTCGAAGATAGCAAAAAGAGCCATGTCCCTCAAAAAGAAATCA 841
QY 888 GCAGGGGGAATGTTTATCCCGTTGAGCCATTTTGGACCCCAAGTACATCTATTCTC 947
DB 840 GCAGGGGGAATGTTTATCCCGTTGAGCCATTTTGGACCCCAAGTACATCTATTCTC 781
QY 948 GAGCGTCACCTGGGACCTCTCACTGACGAAGATACGAGTACGATGAGTACGATACCTA 1007
DB 780 GAGCGTCACCTGGGACCTCTCACTGACGAAGATACGAGTACGATGAGTACGATACCTA 721
QY 1008 TCATTTGCTGTAAATGATATATCCAGAGTCTTAATATGTGATGATGTTTCCGTGCA 1067
DB 720 TCATTTGCTGTAAATGATATATCCAGAGTCTTAATATGTGATGATGTTTCCGTGCA 661
QY 1068 AGGGGATACATATCCAGACCTTGGCCGTAGACATGCTGAAACCAAGGCATTACCGC 1127
DB 660 AGGGGATACATATCCAGACCTTGGCCGTAGACATGCTGAAACCAAGGCATTACCGC 601
QY 1128 ATTACACAGTTTATACCTGCAACAGATGATGCTGACAAATTGGGCGACAACTTTAC 1187
DB 600 ATTACACAGTTTATACCTGCAACAGATGATGCTGACAAATTGGGCGACAACTTTAC 541
QY 1188 AACTCGTAGATGTGATGAGTCCATGATCTTACTCATTTGCAATTTTGTGAAGAGAA 1247
DB 540 AACTCGTAGATGTGATGAGTCCATGATCTTACTCATTTGCAATTTTGTGAAGAGAA 481
QY 1248 CTGATGCTGATTAAGATTGCGGTGAACGCTGCTGTAGAAAGAGATGCTGACATTTGCT 1307
DB 480 CTGATGCTGATTAAGATTGCGGTGAACGCTGCTGTAGAAAGAGATGCTGACATTTGCT 421
QY 1308 AGTATTTTCAGGGGCTAAAGCTGTGACGTATGATACACACAAATTTCTTGGACGTTACT 1367
DB 420 AGTATTTTCAGGGGCTAAAGCTGTGACGTATGATACACACAAATTTCTTGGACGTTACT 361
QY 1368 GGGGATCTAGACAAGATGTTGCACTGCAAGGTATTTGGAGCCCTATGATATGTAG 1427
DB 360 GGGGATCTAGACAAGATGTTGCACTGCAAGGTATTTGGAGCCCTATGATATGTAG 301
QY 1428 GTTGCAAGAACGGTGTGTGCAATTTGCTGCAATGGGAGTGGACATCAAGTACCTT 1487
DB 300 GTTGCAAGAACGGTGTGTGCAATTTGCTGCAATGGGAGTGGACATCAAGTACCTT 241

Oy	1488	CGGGATACATCCCTTTCTTTTAACAGCGTAACCCGTTGAGAGTGATCATCGAACATCA	1547
Db	240	CGTGATACTCTCTTTCTTTTAACAGCGTAACCCGTTGAGAGTGATCATCGAACATCA	181
Oy	1548	GAACCTTTGGAAAGTAAGTTTCATTACACAGTCTATGA--ACCTCAAAACAGACAG	1604
Db	180	GAACCTTTTGAAGTAAGTTTCATTACACAGTCTATGAGGCAACTGAAGACAGACAG	121
Oy	1605	AGAGACTCGCTCGATATATGTTTGTGACTTTGTTTATGAACAATTAAGCTGATTTTGGGC	1664
Db	120	AGAGACTCGCTGATATATGTTTGTGACTTTGTTTATGAACAATTAAGCTGATTTTGGCT	61
Oy	1665	TTTCATT 1670	
Db	60	TCATT 55	
RESULT 3			
ID	ABA98820	standard; DNA; 1721 BP.	
XX	ABA98820;		
DT	18-JUN-2002	(first entry)	
XX			
DE	Soybean ALS small subunit (clone sdc2c.pk001.b10) coding sequence.		
XX			
KW	ALS: Acetolactate synthase; acetohydroxy synthase;		
KW	amino acid biosynthesis; plant; herbicide; ALS small subunit;		
KW	ALS holoenzyme; crop protection chemical; expressed sequence tag;		
KW	EST; ss.		
XX			
OS	Glycine max.		
XX			
PH	Key	Location/Qualifiers	
FT	CDS	88..1433	
FT		/*tag= a	
FT		/product= "ALS small subunit"	
FT		/EC_number= "4.1.3.18"	
XX			
PN	US2001044939-A1.		
XX			
PD	22-NOV-2001.		
XX			
PF	08-DEC-2000; 2000US-0732618.		
XX			
PR	04-JAN-2000; 2000US-174437P.		
XX			
PA	(ABELL/) ABELL L M.		
PA	(FALCO/) FALCO S C.		
XX	(FAMO/) FAMODU O O.		
P1	Abell LM, Falco SC, Famodu OO;		
XX			
DR	WP1: 2002-113395/15.		
XX	P-PSDB: ABB08416.		
PT	New small subunits of plant acetolactate synthase and nucleic acids		
XX	encoding them, useful in screening for novel crop protection chemicals		
PT	or potential herbicidal compounds based upon holoenzyme inhibition		
XX			
PS	Claim 7; Page 37; 57p; English.		
XX			
CC	The invention relates to new isolated polypeptides, which comprises an		
CC	acetolactate synthase (ALS) small subunit. Acetolactate synthase is the		
CC	first committed step in branched chain amino acid biosynthesis in plant		
CC	and bacteria. The polypeptides are useful in aiding the discovery of ne		
CC	herbicides that inhibit plant acetolactate synthase activity. The		
CC	polypeptides are useful for preparing plant ALS holoenzyme, which are		
CC	useful in screening for potential herbicidal compounds based upon		
CC	holoenzyme inhibition. The ALS holoenzyme is also useful in screening		
CC	for novel crop protection chemicals. Particularly, the polypeptides can		
CC	be used as a target to facilitate design and/or identification of		

CC	inhibitors of those enzymes that may be useful as herbicides. Thus, the
CC	polypeptides could be appropriate for new herbicide discovery and
CC	design. The polynucleotides are also useful for producing the
CC	polypeptides and for screening for new herbicides. The current sequence
CC	represents the soybean ALS small subunit (clone pSdc2c.pk001.b10) coding
CC	sequence.
XX	
XX	
50	Sequence 1721 BP: 472 A; 401 C; 391 G; 457 T; 0 other:
Query Match	41.9%; Score 701.8; DB 24; Length 1721;
Best Local Similarity	71.3%; Pred. No. 3.1e+207;
Matches	925; Conservative 0; Mismatches 372; Indels 0; Gaps 0;
OY	285 TCGAGGTGAGAGACACACAAATTTTCAGTATTGTTGGAGACGAAAGCGGAAATTAAT 344
Db	292 TCAAGGTTTCGTCGACACACGATTTCCGTTCGTGCGGAGAGACGGGATGATTAAC 351
OY	345 AGGATTGGACGAGTCTTTGCAAGGAGAGATTCATATTGAAAGCTTCCTTGTCG 404
Db	352 CGGATTTGGCGGGTTCCTGCTAGGAGAGATTAACAATCGAATTCCTCGGGTTGGCC 411
OY	405 AACACAGACAAGGCTCTATTACCATAGTTCGTGTGAAGTGAAGAGGTAAGTTCAGCAG 464
Db	412 AATGAGGACAGGGGCGCTCTTCAACCATCGTCGTGTAGGAGCCGATTAAGTCTGGCCAA 471
OY	465 GTCATCGACCACTTCAGAAAGCTGCTTAATGTTCTAAAGTTGAAGATATCTCAAGTGA 524
Db	472 GTCATGAGACAGCTTCAGAAACTTCGTCATATGCTTAAAGTGAAGATCTTTGAGAGGA 531
OY	525 CCGCAAGTGGAGCGGTGACGTATGCTTTGAAAGTGAATGCATTCACATTCAGAGGCA 584
Db	532 CCACAGGTGGAAACGTGAAGCTGATGCTCAATAAAGTGCATGCGGATCCGAAACCATG 591
OY	585 GAGATCATGTGCTCTGTTGACACATTCAGACAGACAGTGTGATATATGCGGACATGCA 644
Db	592 GAGTTGAAGTGTGTTGCTGAGACATCTTCAGAGCTAAGATTGTGGAATATCTCGGAACAT 651
OY	645 TTGACTATCGAGGTAACTGAGAGATCCTGGAAAAATGATTTGCTGTAAGAAATTTGAAA 704
Db	652 GTGACAATTTGAAGTAACTGAGAGATCCAGGAAAGATGGCGCGTTCAAAAGAAATTTCCGC 711
OY	705 AAGTTTCAGATCAGAGAGATTGTAAAGCAGAGAAAGATAGCACTGAGAAAGGAAAGATG 764
Db	712 AAGTTTGAATTAAGAAATATAGCCAGAACTGGAAGATTTGATTAAGAGGAAAGATG 771
OY	765 GGTGCACTGCTCCATTTTGGCGATTTTCAGAGATCCTATCCAGATCTCAAGAGCAA 824
Db	772 GGTGATCTGCTCCATTTTGGCGGATTTTCAGCTGCTTTATTCACAGATCTTGAAGAGAA 831
OY	835 GCGCCTGTTAGTGTCTTTCGAAGTAGCAAAAAAGAGAGCATTTGCTCCCAAGAGAAACA 884
Db	832 ACACCTGTTAATGCCCTTGTGGAGACAAAATAATGAACCTGTGCCAACTTGATACA 891
OY	885 TCACGAGGGGAGAGATGTTATCCCGTTGAGCCATTTTGGACCCCAAGGTACATCTGAT 944
Db	892 CCTGTGGGGGAGAGTGTATCTTAAGAACCATCAGATGTTTTCAGTCAAGTCAAGTT 951
OY	945 CTCGACGCTCACTGGGAGCTTCTCACTACGCAAGATACGAGTGAAGTACGTCGATCACT 1004
Db	952 CTTCAGCTCACTGGGGTCTCTCAATATGTAAGATACCAAGTGAATTCGATCAACACAG 1011
OY	1005 CTATATGCTGTTAAATGATATTCTACAGAGTCTTAAATATGTACAGTGGTTTTCGCT 1064
Db	1012 TTATCATGCTTTGTGTAACGATGCTCTCTGAGTCTTAAACATTTGTTACAGAGATTTT 1071
OY	1065 CGAAGGGATCAATATTCAGAGCTTTGGCCGTAGAGACATGCTGAACCAAGGCGCATTTCA 1124
Db	1072 AGAAGAGGCTATTAACATTCAGAGTGTACCTGTAGACATGCAAGAGTTGAAGACATTTCT 1131
OY	1125 CGCATTTACACAGTTATATCCTTCGAACAGATGATGGTTCAGCAAAATTTGGTCAGCACT 1184
Db	1132 CGACTTACAACTGTGTGTTCTGTGGAGACAGATGATCAATTTAGCAAGTGTGTGACGACACT 1191

CC plasmid vector comprising the ssuALS nucleotide operably linked to a
CC regulatory sequence can be used to transform host cells for the
CC recombinant production of the protein which is used for evaluating a
CC compound for acetoacetate synthase inhibition and so for selecting
CC potential herbicides. Evaluation of a compound for acetoacetate synthase
CC (ALS) inhibition comprises expression and purification of plant ssuALS
CC from the transformed host, mixing this ssuALS with the large subunit of
CC ALS to form a holoenzyme which is then treated with a test compound.
CC Treated and untreated holoenzyme activity are compared to select
CC compounds with potential for herbicidal activity. Previously, large
CC subunits of plant ALS (*Lsu*ALS) have been isolated, but the existence of
CC ssuALS had not been verified. Mixing the two subunits results in a
CC holoenzyme that has a 4-15 fold increase in specific activity over
CC *Lsu*ALS alone.

XX
XX Sequence 1861 BP; 529 A; 348 C; 432 G; 552 T; 0 other;
S0

Query Match 39.8%; Score 665.8; DB 19; Length 1861;
Best Local Similarity 72.6%; Pred. No. 5.1e-196;
Matches 887; Conservative 0; Mismatches 332; Indels 2; Gaps 2;

OY 288 AAGCGAGCAACACACATTTCTACTATTGTCGAGAAGAACGGC--AATGATTAAATG 346
 ||| | | | | | | |
Db 240 AAGACTAGGCCCATACGCATTTCAGTGTTTGTGTGATGAAGAGGAAATATCAATCG 299

OY 347 GATTCCAGAGTGTTTGCAGCAGACGATACAATAATTGAGACTCTTGCTGTGCTCGAA 406

QY 4

D'b 3

Ov 4

Dib 4

50Y

Db 4

QY 5

Db 5

6 QY

Db 6

QY 7

Dib 6

QY 7

D^b 7

QY 8

Db 7

8 QY

Db 8

9 QY

Db 9

QY 10

Db 9

QY 1067 AAGGGATACAAATATCCAGAGCTTGGCCGTAGACATGCTGAACCAAGGCAATTTTCACG 1126
1020 ACGAGGGTATACAAATCCAAAGTTAGCTTGGACATGCTGAAGTTGAGGGCTTTCTCG 1079
QY 1127 CATTAACAGCTTATACCTGACAGATGGTGGTGGCAAAATTTGGTGAGCAACTTGA 1186
1080 TATTACAGCGTTGTTCTTCCGACAGATGAGTGTAGTGAAGTAGTGGTGCACAACTATA 1139
QY 1187 CAACCTGTAGATGTGCATGAGTGCATGATCTTACTCATTTGGCAATTTCTGAAAGAGA 1246
1140 TAAGTTGGTTGATATTCATGAGTGGGATATTACCTCACCCTCCATTTCCGGAAGAGA 1199
QY 1247 ACATGCTCAATTAAGATTGCTCCGTAAGAGCTGCTGTAGAAAGAGTCTTGACATTCG 1306
1200 ACTAATGTTATTAAGATTGCTGTAAGTCTGACAGCCGCCGCAATGTTCTTGTGATTCG 1259
QY 1307 TAGATTTTGAAGGCTTAAGCTGTGACGTATGCTGATCACAATTTCTTGGACGTTAC 1366
1260 CAGCATTTGAGACCAAAAGCTGTGATGCTGACACCACTATTAATCTTGGAGCTTAC 1319
QY 1367 TGGGATCTAGACAGATGGTGTGACTGCAAAAGTTATTTGAGCCCTATGTTATGTTGA 1426
1320 AGGAGATTGTCATTAAGATGGTGTGCTTGGACGCGCTACTAGAGCTTATGTTATTTGTA 1379
QY 1427 GGTTCGCAAGACCGGTGCTGTGCTGCTGCTGATTCGGAGTGGAGTCCAAATGACT 1486
1380 GGTAGCGCGAA-CAGACGCTGCGCTGCTGCTGCTGATTCAGTGTGATTCGAAATGCT 1438
QY 1487 TCGTGATCTGCTTCTTTT 1507
Db 1439 GCGAGATATTCAATCCCTTT 1459

RESULT 5
ABA98819
ID ABA98819 standard; DNA: 1435 BP.
XX ABA98819:
XX AC
XX DT 18-JUN-2002 (first entry)
XX DE
XX DE Rice ALS small subunit (clone r10n.pk117.a16) coding sequence.
XX AC
XX KW ALS: Acetolactate synthase; acetoaldehyde synthase;
KW amino acid biosynthesis; plant; herbicide; ALS small subunit;
KW ALS holoenzyme; crop protection chemical; expressed sequence tag;
KW EST; ss.
XX OS
XX OS Oryza sativa.
XX FH
XX FH Key Location/Qualifiers
FT CDS 45..1142
FT FT /*tag= a
FT FT /product= "ALS small subunit"
FT FT /EC_number= "4.1.3.18"
XX US2001044939-A1.
XX PN
XX PD 22-NOV-2001.
XX PF 08-DEC-2000; 2000US-0732618.
XX PR 04-JAN-2000; 2000US-174437P.
XX PA (ABELL/) ABELL L M.
XX PA (FALC/) FALCO S C.
XX PA (FAMO/) FAMODU O O.
XX PI Abell LM, Falco SC, Famodu OO;
XX DR WPI: 2002-113395/15.
XX DR P-PSDB: ABB08415.

PT New small subunits of plant acetolactate synthase and nucleic acids
PT encoding them, useful in screening for novel crop protection chemicals
PT or potential herbicidal compounds based upon holoenzyme inhibition
PS Claim 7; page 35; 57pp; English.
XX
XX The invention relates to new isolated polypeptides, which comprises an
CC acetolactate synthase (ALS) small subunit. Acetolactate synthase is the
CC first committed step in branched chain amino acid biosynthesis in plants
CC and bacteria. The polypeptides are useful in aiding the discovery of new
CC herbicides that inhibit plant acetolactate synthase activity. The
CC polypeptides are useful for preparing plant ALS holoenzyme, which are
CC useful in screening for potential herbicidal compounds based upon
CC holoenzyme inhibition. The ALS holoenzyme is also useful in screening
CC for novel crop protection chemicals. Particularly, the polypeptides can
CC be used as a target to facilitate design and/or identification of
CC inhibitors of those enzymes that may be useful as herbicides. Thus, the
CC polypeptides could be appropriate for new herbicide discovery and
CC design. The polynucleotides are also useful for producing the
CC polypeptides and for screening for new herbicides. The current sequence
CC represents the rice ALS small subunit (clone r10n.pk117.a16) coding
CC sequence.

Sequence 1435 BP; 413 A; 282 C; 332 G; 408 T; 0 other;

Query Match 35.5%; Score 593.8; DB 24; Length 1435;
Best Local Similarity 71.3%; Pred. No. 1.1e-173;
Matches 813; Conservative 0; Mismatches 322; Indels 6; Gaps 2;

QY 394 CTGTTGGTCTGAACAGACAGCAAGCTCTATTCACCATGTTGCTGTGCAACTGAAGG 453
19 CCGTGGGGCTCAACAGAGCAAGGCAAGCTATGCTCACTGCTCCGGACGCAAGG 78
QY 454 TACTTCAGAGCTATGAGCACTCAAGAGCTCGTTAATGTTCTTAAGGTTGAAGATA 513
79 TGTCTCAACCAAGTCATCGAGCAGCTCAACAGCTTGTCAACGCTTGAATGTGAAAGATC 138
QY 514 TCTCAAGTGAGCGCAAGTGAGAGCTGATGCTGTAAAGTGAATGACATTCAG 573
139 TATCTAAGAGGCCACAGGTTGAAGAGAGCTGATGCTTAATAAATTAATGTTGAACCA 198
QY 574 AATCCAGGCGAGATCATGTGCTAGTGTACACATTCAGAGCAAGAGTTGATATAG 633
199 ATCAGGCTCTGAGGCTATGCTTTAGTGTATTTTCCAGCAAAATGTTGATATTT 258
QY 634 CGAATCATGCTATGATGATGAGTGAAGTGAAGTCTGGAATGCTGTTGA 693
259 CGGAGAACCTTACCATGAGGTAACTGGAGATCTGGCAAAATTTGCTGTCAAA 318
QY 694 GAAATTTGAAAAAGTTTCAGATCGAGAGATGTAAGGACAGAGAAATAGCACTGAGAA 753
319 GGAACCTCAGCAAAATTTGGATTAAGAAATTTGTGAACGGGAAAAATTTGCTTGAGAC 378
QY 754 GGGAAAAAGATGGTGCATGCTGCTATTTGGCGATTTTTCAGCAGATCTATTCAGATC 813
379 GTGAAAAAATTTGGAGCAAGTCCCGCTTCCGGGATTTTCTGCTGCTTACCCAGATC 438
QY 814 TCAAGGAGCAAGCGCTGTAGTGTCTTCGAA--GTAGCAAAAAAGAGCCATGTGC 870
439 TCATAGAGCATTTGCCAAAAAATTTCTTCTTACTTGTAAATAAGACGTAAATGAGAA 498
QY 871 CTCAAAAAGAAACATC--AGCAGGGGAGATGTTATCCCGTTGAGCATTTTTCAGC 927
499 GTTTTATCAACCATTCATGCTGCGGGGAGATGCTATCTGTGGAACCTTATGAGAGTT 558
QY 928 CCAAGGTACATGCTATTCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 987
559 CATCCATGAACCAAGTACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 618
QY 988 GACTAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1047
619 GACTTGCATCATTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 678

Qy	1048	TCAGTGTGTTTTCGCTCCGAAGGGGATACAAATATCCAGAGCTGGCCGTAGSACATGCTG	1107
Db	679	TTACAGGGGCTTTGGCTTCGAGAGGCTCAAAATATACAGAGCTTCTTAGGGCCACGCTG	738
Qy	1108	AAACCAAGGCGCATTTTACACGATATACAAAGATTATACCTGCATCAACATGATCAATCGGCACCA	1167
Db	739	AAAAGTCAGGCCCTTTCCGCTATTACAAACGTTGCTTCGAGACAGATATATCATTGACGA	798
Qy	1168	AATTGTGTCAGCAACTTTACAAATCTGATAGATGTGCATGAGGTCCATGATCTTAATCATY	1227
Db	799	AGTTAGTTCAAGCAAGCTTAACAAACTTTGTGATGTGATGAGTTCAAGATATAACTCACT	858
Qy	1228	TGCCATTTTTCGAAGAAGAACTATGCTGATTAACATTTCCTCGTAACCGTGTCTGTAGA	1287
Db	859	TGCCATTTCGTGAAGAAGAACTATGCTTATCAAGGTTTCTGTGAACACTGCTGTCTGGA	918
Qy	1288	GAGATGTCCTGGACATATTCAGTAGTAATTTTCAGGGCTTAACCTGTGACGATCTGATCACA	1347
Db	919	GAGCACTACTAGATATATTTGCTGAAATCTTCCGGCCAAATCTGTTATGTTCTGTGATCACA	978
Qy	1348	CAATTGATTTGCAGCTTACTGGGGATCTAGACAAAGATGGTTCACCTGCAAAAGTTATTTGG	1407
Db	979	CTGTATGATTTACAGCTTACTGGGGATCTGACAAAGATGTTTCATTTACAAAGCGCTTTGG	1038
Qy	1408	AGCCCTATGCTATATATGTGAGGTTGCAGAAACCGGTCGTGTGSCATTTGGCTGTGAATCGG	1467
Db	1039	AGCCTTATATGCAATCTGTGAGGTCGCCAAGAACGGCGGATGGTCCGCGCAATCCG	1098
Qy	1468	GAGTGGACTCCAGTACTCTTCGCGGATCTGATCCCTTCTTTTAAAGCGCTAAACCGTTTGAG	1527
Db	1099	GTGTGCAATCCAAAGTACTCTTCGAGCTACTCTTCTGTTGAATCCAGAGCTTTTGAG	1158
Qy	1528	A 1528	.
Db	1159	A 1159	
RESULT 6			
ABA98816			
ID	ABA98816	standard; DNA; 1813 BP.	
XX	ABA98816;		
MC			
XX			
DT	18-JUN-2002	(first entry)	
XX			
DE	Corn ALS small subunit (clone p0094.cssl172ra) coding sequence.		
XX			
KW	ALS: Acetolactate synthase; acetoaldehyde synthase;		
KW	amino acid biosynthesis; plant; herbicide; ALS small subunit;		
KW	ALS holoenzyme; crop protection chemical; ss.		
XX			
OS	Zea mays.		
XX			
TH	key	Location/Qualifiers	
FT	CDS	1..1452	
FT		/*lag= a	
FT		/partial	
FT		/product= "ALS small subunit"	
FT		/Ec_number= "4.1.3.18"	
FT		/note= "no start codon present"	
XX			
PN	US2001044939-A1.		
XX			
PD	22-NOV-2001.		
XX			
PF	08-DEC-2000; 2000US-0732618.		
XX			
PR	04-JAN-2000; 2000US-174437P.		
XX			
PA	(ABEL/) ABELL L M.		
PA	(FALC/) FALCO S C.		
PA	(FAMOD/) FAMODU O O.		
XX			

PI Abell LM, Falco SC, Farnodu OO;
XX
DR WPI: 2002-113395/15.
XX P-PSDB; ABB08412.
XX
XX New small subunits of plant acetoacetate synthase and nucleic acids
PT encoding them, useful in screening for novel crop protection chemicals
or potential herbicidal compounds based upon holoenzyme inhibition -
PS
XX
XX Claim 7, Page 31: 57pp: English.
XX
XX The invention relates to new isolated polypeptides, which comprises an
CC acetoacetate synthase (ALS) small subunit. Acetoacetate synthase is the
CC first committed step in branched chain amino acid biosynthesis in plant
CC and bacteria. The polypeptides are useful in aiding the discovery of new
CC herbicides that inhibit plant acetoacetate synthase activity. The
CC polypeptides are useful for preparing plant ALS holoenzyme, which are
CC useful in screening for potential herbicidal compounds based upon
CC holoenzyme inhibition. The ALS holoenzyme is also useful in screening
CC for novel crop protection chemicals. Particularly, the polypeptides can
CC be used as a target to facilitate design and/or identification of
CC inhibitors of those enzymes that may be useful as herbicides. Thus, the
CC polypeptides could be appropriate for new herbicide discovery and
CC design. The polynucleotides are also useful for producing the
CC polypeptides and for screening for new herbicides. The current sequence
CC represents the corn ALS small subunit (clone p0094.ces1121a) coding
CC sequence.
XX
XX Sequence 1813 BP; 545 A; 370 C; 432 G; 466 T; 0 other;

Query Match	35.3%	Score 590;	DB 24;	Length 1813;
Best Local Similarity	68.2%;	Pred. No. 1.9e-172;		
Matches 834;	Conservative 0;	Mismatches 385;	Indels 3;	Gaps 1

OY	285	TCGAAGTGAGAAACACACAAATTCAGTATTGTTGGAGACGAAACCGAATGATTAAT	344
Db	223	TCGGTTATGAAGCGCTCACACCCCTTACAGTTTTTTATGTGTGTGAAGTGGGATGATCAAT	282
OY	345	AGGATTGCAGAGAGTGTTCGAAGGAGAGGATACAAATATTGAGATCTTGGCTGGTGCTG	404
Db	283	CGAATTGCTGGGGTTTTTGGCTAGAGAGAGATATTAACATCGAGTCATTGGCTGTGGGTTG	342
OY	405	AACAGAGACAAGGCTCTATTCCACCATAGTTGCTGTGGAACGTAAGGATCTCAGACG	464
Db	343	AACAGAGATTAAGCATTTATTACATATAGTACTGTGCAGAAACAGACAATAATATTAAACG	402
OY	465	GTGATCGAGCAATCTCAGAACCTGTTATGTTTAAAGGTGAAGTATCTCAAGTGAG	524
Db	403	GTCCGTAAGCAACTTAACAAACCTGTGTAATGCTATAAAGGTGATGATTTATCAATGAA	462
OY	525	CCGCAAGGAGACCGGAGCCTGATCCTGTAAACTGTAATGCACATCCAGATCCAGGCA	584
Db	463	CCACAGTTTGAAAGAGAACTTATGCTTTATTAAGTTAAATGACAGCGCGAAGAACTTACCT	522
OY	585	GAGATCATGTGGCTAGTTGACACATTCAGAGCAAGAGTTGATATATAGCGGAACATGCA	644
Db	523	GAGATTAATGGGTGGTTCGGCATTTTCAACACAGAAAGCGTTGATCTTTCAGACTACACA	582
OY	645	TTGACTATCGAGGTAAGTGGAGATCTCGGAAATAATGATTCGTGAGAGAAATTTGAAA	704
Db	583	CTAATCTATTGGAGTAACTCGAGATCTCGGAAGATGTTGTCGAATATACGAAGACTCTAGC	642
OY	705	AAGTTTCAGATCAGAGATTTGTAAGGACAGAGAAAGATAGCACTGAGAGAAAGAAATG	764
Db	643	AAATGTGGGATGAGGAATTTGCTAGAACTGCGCAAGATAGCTTTGGCCCGGAAAAAATG	702
OY	765	GGTGCACATGCTCCATTTTGGCGATTTTTCACACACATCTTACAGATCTCAAGAGACAA	824
Db	703	GGAGAAACTGCTCCATTTTGGAGGTTCTCTGACACTTCTTATCCGGATCTTCGAAGTGGCA	762
OY	825	GGCGCTGTACTGTTCTTGAAGTAGCAAAAAAGAGCATTTGTCCTCAAAAGCAACA	884
Db	763	ATACCTTCAAAATTTCCAGCAAAACAATGTTGTAAGCAATCATCAAGAAATCCGAGCAAA	822

```

QY 885 TCAGAGGGGAGATGTTATCCCGTTGAGCCATTT---TTTGACCCCAAGGTACATCGT 941
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 823 TCTTCAGGGGGATGATTTATCCAGTAGATCTTATGAAGGCTTCTCATCAACAGTCAA 882
QY 942 ATTTCGACGCTTACCTGGGAGCTTCTCACTGACGAAGATACGATGAGTACGTCGCT 1001
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 883 ATTTCGATGCTATTTGGGCTTTATGACTGATGGCGATCCAAACAGGTTTGTTCACAT 942
QY 1002 ACTCTATCATTTGCTTAAATGATATTTCCAGAGATCTTAAATATTTGATGCTGTTTC 1061
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 943 ACTCTATCATTTCTTGTGATGATGTCCTCGAGTTCCTCAATCTTGTAAACAGTATTTC 1002
QY 1062 GCTCGAAGGGATACAAATTCAGAGCTTGGCGTAGACATGCTGAACCAAGGCGATT 1121
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1003 TCCAGAAAGGGCTACAAATATAGAGCTTTGGCTTTGGCCAGCTGAAGAAAGAGAACT 1062
QY 1122 TCACGATTCACACAGTATTACCTGCAACAGATGAATCGGTGACGAATTTGTCAGCAA 1181
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1063 TCTCGCATACACTGTTGTTCTGGAAGTGAATTCATTCGCAAGCTAGTACATCAA 1122
QY 1182 CTTTACAACGCTAGATGTCATGATGTCATGATCTTACTATTTGCCATTTTCTGAA 1241
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1123 CTGTACACACTATTTATGATTTATGAGTTTCAGAGATTTTACCACCTTACCATTGCTGCT 1182
QY 1242 AGAGAACTGATGCTGATTAAGATTTGCCGTAACGCTGCTGTAAGAGAGATGCTTGAC 1301
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1183 AGAGATTAATGATCAATAAAGTGGCGCAAAATGCTACAGCTGAAGGAGATGCTTAGAT 1242
QY 1302 ATTGCTAGATTTTTCAGGGCTAAAGCTGTTGACGATCTGATCAGCAATTACTTTGCAG 1361
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1243 ATTGCTACATTTTTCAGGACAGAAAGTTCATATTCAGACACCAATTTACACTACTG 1302
QY 1362 CTTACTGGGATCTACAGAAATGTTGCACTGCAAGGTTATTGGAGCCATGATGTATA 1421
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1303 CTCACCGGAGACATTGACGAATGTTAGATTGCCAAAAGATGCTAGAGCAGATGAGCATC 1362
QY 1422 TGTGAGTTGCAAGAACCGGTGCTGTGCGATTTGGCTGGAATCGGAGTGCATCCAG 1481
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1363 TGTGAGTTGCAAGAACCGGATTTGCTGTGCTCCGAGAGTGTGAGTTGACTCCAG 1422
QY 1482 TACCTTCGTGATCTCTTTTC 1503
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1423 TACCTCCGCGGTTTCCCTCC 1444
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7
ABA98814
ID ABA98814 standard; DNA: 1297 BP.
AC ABA98814;
XX
XX
DT 18-JUN-2002 (first entry)
XX
DE Corn ALS small subunit (clone cen3n.pk0112.c11) coding sequence.
XX
KM ALS: Acetolactate synthase; acetoaldehyde synthase;
XX amino acid biosynthesis; plant; herbicide; ALS small subunit;
XX ALS holoenzyme; crop protection chemical; ss.
XX
OS Zea mays.
XX
XX
XX Key Location/Qualifiers
XX CDS 2..1054
XX FT /tag=a
XX FT /partial
XX FT /product="ALS small subunit"
XX FT /EC_number="4.1.3.18"
XX FT /note="no start codon present"
XX
XX US2001044939-A1.
XX
XX 22-NOV-2001.

```

```

XX
XX 08-DEC-2000; 2000US-0732618.
XX
XX
XX
XX 04-JAN-2000; 2000US-174437P.
XX
XX (ABELL/) ABELL L M.
XX (FALCO/) FALCO S C.
XX (FAMO/) FAMODU O O.
XX
XX
XX Abell LM, Falco SC, Famodu OO;
XX
XX WPI: 2002-113395/15.
XX P-PSDB: ABB08410.
XX
XX New small subunits of plant acetolactate synthase and nucleic acids
XX encoding them, useful in screening for novel crop protection chemicals
XX or potential herbicidal compounds based upon holoenzyme inhibition -
XX
XX Claim 7, Page 28; 57pp; English.
XX
XX The invention relates to new isolated polypeptides, which comprises an
XX acetolactate synthase (ALS) small subunit. Acetolactate synthase is the
XX first committed step in branched chain amino acid biosynthesis in plants
XX and bacteria. The polypeptides are useful in aiding the discovery of new
XX herbicides that inhibit plant acetolactate synthase activity. The
XX polypeptides are useful for preparing plant ALS holoenzyme, which are
XX useful in screening for potential herbicidal compounds based upon
XX holoenzyme inhibition. The ALS holoenzyme is also useful in screening
XX for novel crop protection chemicals. Particularly, the polypeptides can
XX be used as a target to facilitate design and/or identification of
XX inhibitors of those enzymes that may be useful as herbicides. Thus, the
XX polypeptides could be appropriate for new herbicide discovery and
XX design. The polynucleotides are also useful for producing the
XX polypeptides and for screening for new herbicides. The current sequence
XX represents the corn ALS small subunit (clone cen3n.pk0112.c11) coding
XX sequence.
XX
XX Sequence 1297 BP; 356 A; 263 C; 303 G; 375 T; 0 other;
XX
XX
XX Query Match 32.1%; Score 537.6; DB 24; Length 1297;
XX Best Local Similarity 69.6%; Pred. No. 3,1e-156;
XX Matches 729; Conservative 0; Mismatches 319; Indels 0; Gaps 0;
XX
XX
XX 460 AGCAGTCATCGCAACCTCCGAGAGCTGTAATGTTCTTAAGTGAAGATATCTCAA 519
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3 ACCAAGTCAATTAGCAGCTCAATAAGCTTCGTAACCTTCATAGTCTATCTA 62
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX
XX 520 GTGACCCGCAAGTGAAGCGTGATGCTTTGTAAGATGTAATCCAGATCCAGATCCA 579
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 63 AAGAACCTCAGGTTGAAGAGAGCTGATGCTTAAGCTTAACGTTGAACCTGATCAGC 122
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX
XX 580 GGGCAGACATCATTTGGCTAGTTGACACATTCAGAGCAAGATTTAGATATAGGGGAA 639
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 123 GCCCTGAGTCAATGTTTATTTAGTTTACATTTTCAGCAAAATGTTTATATATCTGAGA 182
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX
XX 640 ATGCAATTGACTATCCAGATTAATCTGAGATCTCGAATAATGATCTCTAGCAAAATTT 699
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 183 AAACCTTACCTTAAGAGTGTGAGATCTCGGCAAAATTTCTCAAGTGCAGAGAGATTC 242
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX
XX 700 TGAAGAAGTTTCAGATTCAGAGAGATTTGTAAGAGCAAGAAAGTACCTAGACAGGAAA 759
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 243 TAAGCAAAATTCGACATCAAGAAATTTGACAGACAGGAAATTTCTTTGAGACGTGAAA 302
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX
XX 760 AGATGGGTGCAACTGCTCATTTTGGCGATTTTTCAGCAGCATCTTATCCAGATCTCAAG 819
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 303 AGATTGGTGCAACACCGCTTTCTGCGATTTTCTGCTTTTATCCAGACCTTAAG 362
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX
XX 820 AGCAAGCCCTGTTAGTCTTTCGAAAGTGAAGAAAGAGCAATTCCTCCCTCAAAAG 879
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 363 AGGCATTACCAAAAACCGCTTACATCTGTAATTAAGACAGTGAATGCACTTTTGTTC 422
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX
XX 880 AAACATCAGCAGGGGAGATGTTTATCCCGTTGAGCCATTTTTCAGCCCAAGGTACATC 939
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```


XX (AMCY) AMERICAN CYANAMID CO.
 XX
 PI Kakefuda G, Costello C, Sun M, Hu W,
 XX
 DR WPI: 2000-365633/31.
 DR P-PSDB: AAY96203.
 XX
 PT New polynucleotide encoding eukaryotic acetylhydroxy-acid synthetase
 PT small subunit protein for producing transgenic herbicide resistant
 PT plants and identifying mutations affecting enzymatic activity of the
 PT synthetase
 XX
 PS Claim 3: Page 53-54; 57pp; English.
 XX
 CC The present sequence is the genomic sequence for Arabidopsis
 CC acetylhydroxy-acid synthase (AHAS) small subunit protein. AHAS is also
 CC known as acetolactate synthase. This enzyme is needed for branched-chain
 CC amino acid synthesis and so is essential for life. Inhibition of this
 CC enzyme would lead to plant death and therefore inhibitors would be
 CC potential herbicides. Certain herbicides are known to inhibit AHASs:
 CC imidazolinones, sulfonureas, triazopyrimidine sulfonamides,
 CC pyrimidyl-oxy-benzoic acids, sulfamoylureas and sulfonycarboxamides.
 CC The present sequence may be used to identify mutant AHASs which are
 CC resistant to these herbicides and may be used to create herbicide
 CC resistant transgenic plants e.g. dicot and monocot crop plants.
 CC
 SQ Sequence 4895 BP; 1296 A; 884 C; 1011 G; 1703 T; 1 other;
 Query Match 25.0%; Score 419; DB 21; Length 4895;
 Best local Similarity 86.7%; Pred. No. 5.4e-119;
 Matches 507; Conservative 0; Mismatches 0; Indels 78; Gaps 1;
 XX
 YY 1 GTCTCTTCAGTGAACAAAACCTTCGGCTTCGTCGCAATGGGGCCATTTCGTAA 60
 DB 717 GTCTTTCAGTGAACAAAACCTTCGGCTTCGTCGCAATGGGGCCATTTCGTAA 776
 YY 61 GTCTTCACCATTCATTCGCTTCGTCGATGCGATGTCGATTCCTTCCTGCTTG 120
 DB 777 GTCTTCACCATTCATTCGCTTCGTCGATGCGATGTCGATTCCTTCCTGCTTG 836
 YY 121 TATCTTCGACGCGTTCATTCCTCCGCGCAAGATTCATATCTCTCCGATATCTTCCG 180
 DB 837 TATCTTCGACGCGTTCATTCCTCCGCGCAAGATTCATATCTCTCCGATATCTTCCG 896
 YY 181 ACCGTGGCGATGAATGGGTAAAGATGAAGATGGAAGATTCGTAAAGCGTCGATGGGAAG 240
 DB 897 ACCGTGGCGATGAATGGGTAAAGATGAAGATGGAAGATTCGTAAAGCGTCGATGGGAAG 956
 YY 241 TCTCTGATGCGTTCCTTCGGAAGCTTCATCTCGACTCCCAAAATCGAA----- 289
 DB 957 TCTCTGATGCGTTCCTTCGGAAGCTTCATCTCGACTCCCAAAATCGAAAGCGACTGTGAA 1016
 YY 290 ----- 289
 DB 1017 TAATATTGCTTAAGTCTTCCTTTTGCCCTTCCTTTGATTGATTCCTTTGTCATTA 1076
 YY 290 -----GGTGAAGAGACACACAAATTCAGTATTGTTGGAGACGAAGCGAATGATTA 342
 DB 1077 AAATCAGAGGAGGAAGACACACAAATTCAGTATTGTTGGAGACGAAGCGAATGATTA 1136
 YY 343 ATAGAGATTGAGAGATGTTGTCGAAGAGAGATACAAATATTGAGAGTCTTCGTTGTCGTC 402
 DB 1137 ATAGAGATTGAGAGATGTTGTCGAAGAGAGATACAAATATTGAGAGTCTTCGTTGTCGTC 1196
 YY 403 TGAACAGAGACGAAGCTTCATTCACCATAGTTCGTGTGAACTGAAAGGATCTTCAGC 462
 DB 1197 TGAACAGAGACGAAGCTTCATTCACCATAGTTCGTGTGAACTGAAAGGATCTTCAGC 1256
 YY 463 AGGTCAATCGAGCAATTCGCAAGACCTGTTAATGTTCTAAAGGTTG 507
 DB 1257 AGGTCAATCGAGCAATTCGCAAGACCTGTTAATGTTCTAAAGGTTG 1301

RESULT 9
 AAV55854
 ID AAV55854 strand; DNA; 611 BP.
 XX
 AC AAV55854;
 XX
 DT 18-NOV-1998 (first entry)
 XX
 DE Small subunit of corn acetolactate synthase insert nucleotide sequence.
 XX
 KW ALS; small subunit; acetolactate synthase; plant; ssuALS; lsuALS;
 KW Nicotiana glauca; plumbaginifolia; plasmid vector; herbicide; holoenzyme;
 KW large subunit; ss.
 XX
 OS ~~See entry~~
 XX
 PN W09837206-A1.
 XX
 PD 27-APR-1998.
 XX
 PF 23-FEB-1998; 98MO-US03506.
 XX
 PR 24-FEB-1997; 97US-0039148.
 XX
 PA (DUPD) DU POINT DE MEMOIRS & CO E I.
 XX
 PI Abell LM, Hershey HP;
 XX
 DR WPI: 1998-467568/40.
 XX
 PT New nucleic acid encoding a plant aceto:lactate synthase small
 PT subunit - that combines with the large subunit to give a
 PT holoenzyme having higher activity than the large subunit alone
 XX
 PS Example 1; Page 29; 47pp; English.
 XX
 CC This represents the nucleotide sequence of the cDNA insert encoding
 CC the small subunit of corn acetolactate synthase found in the plasmid
 CC clone m5.12.b12.sk20. This is used for the isolation of the cDNA
 CC encoding the small subunit of a plant acetolactate synthase (ssuALS). A
 CC plasmid vector comprising the ssuALS nucleotide operably linked to a
 CC regulatory sequence can be used to transform host cells for the
 CC recombinant production of the protein which is used for evaluating a
 CC compound for acetolactate synthase inhibition and so for selecting
 CC potential herbicides. Evaluation of a compound for acetolactate synthase
 CC (ALS) inhibition comprises expression and purification of plant ssuALS
 CC from the transformed host, mixing this ssuALS with the large subunit of
 CC ALS to form a holoenzyme which is then treated with a test compound.
 CC treated and untreated holoenzyme activity are compared to select
 CC compounds with potential for herbicidal activity. Previously, large
 CC subunits of plant ALS (lsuALS) have been isolated, but the existence of
 CC ssuALS had not been verified. Mixing the two subunits results in a
 CC holoenzyme that has a 4-15 fold increase in specific activity over
 CC lsuALS alone.
 CC
 SQ Sequence 611 BP; 158 A; 133 C; 139 G; 175 T; 6 other;
 Query Match 14.5%; Score 243.4; DB 19; Length 611;
 Best local Similarity 70.7%; Pred. No. 6.6e-65;
 Matches 378; Conservative 0; Mismatches 150; Indels 7; Gaps 4;
 XX
 YY 888 GCAGGGGAGATGTTATCCGTCGACCATTTTGGACCCCAAGGTACATGTAATTC 947
 DB 77 GCTGGGGGAGATGTTATCCGTCGACCATTTTGGACCCCAAGGTACATGTAATTC 136
 YY 948 GACGCTCACTGGGAGATTCCTCACTGACGAGATAGAGTGCAGTGCAGTACTCTTA 1007
 DB 137 GATGCTCACTGGGAGATTCCTCACTGACGAGATAGAGTGCAGTGCAGTACTCTTC 1296
 YY 1008 TCATTCCTTTGAATGATATTCAGAGATGTTTAAATATGTCAGTGGTTCCTCGCA 1067
 DB 197 TCATTCCTTTGAATGATATTCAGAGATGTTTAAATATGTCAGTGGTTCCTCGC 256

FT	/partial
FT	/product= "ALS small subunit"
FT	/EC_number= "4.1.3.18"
FT	/transl_except= (pos:105..107, aa:Xaa)
FT	/transl_except= (pos:129..131, aa:Xaa)
FT	/transl_except= (pos:165..167, aa:Xaa)
FT	/transl_except= (pos:171..173, aa:Xaa)
FT	/transl_except= (pos:174..176, aa:Xaa)
FT	/transl_except= (pos:192..194, aa:Xaa)
FT	/transl_except= (pos:198..200, aa:Xaa)
FT	/transl_except= (pos:357..359, aa:Xaa)
FT	/transl_except= (pos:390..392, aa:Xaa)
FT	/transl_except= (pos:447..449, aa:Xaa)
FT	/transl_except= (pos:453..455, aa:Xaa)
FT	/note= "Xaa is unknown; no start or stop codon present"
PN	US2001044939-A1.
XX	
PD	22-NOV-2001.
XX	
PE	08-DEC-2000; 2000US-0732618.
XX	
PR	04-JAN-2000; 2000US-174437P.
XX	
PA	(ABELL) ABELL L M.
PA	(FALCO/) FALCO S C.
PA	(FAMODU/) FAMODU O O.
P1	
PI	Abell LM, Falco SC, Famodu OO;
XX	
DR	WPI: 2002-113395/15.
DR	P-PSDB: ABB08409.
PT	New small subunits of plant acetoacetate synthase and nucleic acids
PT	encoding them, useful in screening for novel crop protection chemicals
PT	or potential herbicidal compounds based upon holoenzyme inhibition -
XX	
PS	Claim 7; page 27; 57pp; English.
XX	
CC	The invention relates to new isolated polypeptides, which comprises an
CC	acetoacetate synthase (ALS) small subunit. Acetoacetate synthase is the
CC	first committed step in branched chain amino acid biosynthesis in plants
CC	and bacteria. The polypeptides are useful in aiding the discovery of new
CC	herbicides that inhibit plant acetoacetate synthase activity. The
CC	polypeptides are useful for preparing plant ALS holoenzyme, which are
CC	useful in screening for potential herbicidal compounds based upon
CC	holoenzyme inhibition. The ALS holoenzyme is also useful in screening
CC	for novel crop protection chemicals. Particularly, the polypeptides can
CC	be used as a target to facilitate design and/or identification of
CC	inhibitors of those enzymes that may be useful as herbicides. Thus, the
CC	polypeptides could be appropriate for new herbicide discovery and
CC	design. The polynucleotides are also useful for producing the
CC	polypeptides and for screening for new herbicides. The current sequence
CC	represents the corn ALS small subunit (clone cen3n.pK0112.c11) EST
CC	fragment.
XX	
SQ	Sequence 470 BP; 146 A; 87 C; 110 G; 110 T; 17 other;
	Query Match 11.8%; Score 197; DB 24; Length 470;
	Best Local Similarity 64.3%; Pred. NO. 1.5e-50;
	Matches 297; Conservative 0; Mismatches 162; Indels 3; Gaps 1,1,1
OY	460 AGCAGGTCATGCAGCAACCTCCGAAGCTCGTTAAAGTGTGAAGTAGATCTCAA 519
Dd	1 ACCAAGTCATTGAGCAGCTCATTAAGCTTCGTCACAGCTTCATAGTGTTGAAGATCTACTA 60
OY	520 GTGAGCCGCAAGTGAGCGTGAGCTGATGCTTTGTAAGAAGTGAATGCACATCCAGATCCA 579
Dd	61 AAGAACCTCAGGTTGAAGAGAGAGCTGATGCTTATAAAGCTAAACNNAGCACTGATCACG 120
OY	580 GGGCAGAAATATGATGCGCTAGTTGACACATTCAGAGCAAGAGTTGTAGATATAGCGGAAC 639
Dd	121 GCCCGAGAGNCATGGTTTTAGTTGACATTGTTGAGGCCAAAAGTNTGNNTGATANANC TGACA 180

Oy	640	ATGCATTGACTATTCAGTAGTAAGTAAGTCTCGAAAAATGATTGCTGTGAGAAAGAAATT	659
Oy	181	AAACACTTACCCANNAAGAGTGGAGATCCTGGCAAAATTGGCTAGTGCAGAGGAATC	240
Oy	700	TGAAAAAGTTTCAGTTCAGAGAGATTTGTAAGACAGAGAAAGATAGCATGAGAGAAAGGAAA	759
Db	241	TAAAGAAATTCGGGACTCAAGAAATTTTCAGGACAGGAAAAATTTGCTTGGACGAGNAAA	300
Oy	760	AGATGGGTGCAACTGCTCCATTTTGGCGATTTTTCAGCAGCATCCTATCCAGATCTCAAG	819
Db	301	AGATTGGNAGCAACAGCCCGTTTCTGGCGATTTTCTGCTGCTTTATCCAGACCTTANAG	360
Oy	820	ACCAAGCCGCTGTGTAGTCTTCTTGCAGACTACCAAAAAAGAGCCATTGTCCCTCAAAAG	879
Db	361	AGCATTTACCAAAAAACCGCTTACATCTGNAATTAAGACAGTGAATGCAATGCAATTTTGTTC	420
Oy	880	AAACATC--AGCAGGGGAGATGTTTATCCCTGGTGAAGCAT	918
Db	421	GACCATCCATGCTGGGGGTGATGTTNATCNGNGAATCTT	462
RESULT 12			
ABA98817			
ID	ABA98817	standard; DNA; 515 BP.	
XX			
AC	ABA98817;		
DT	18-JUN-2002	(first entry)	
XX			
DE			
XX			
KM	ALS: Acetolactate synthase; acetylhydroxy synthase;		
KM	amino acid biosynthesis; plant; herbicide; ALS small subunit;		
KW	ALS holoenzyme; crop protection chemical; expressed sequence tag;		
KM	EST; ss.		
XX			
OS	Oryza sativa.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	1..471	
FT		/*tag= a	
FT		/partial	
FT		/product= "ALS small subunit"	
FT		/EC_number= "4.1.3.18"	
FT		/note= "no start codon present"	
XX			
PN	US2001044939-A1.		
XX			
XX	22-NOV-2001.		
PD			
XX	08-DEC-2000; 2000US-0732618.		
PF			
XX			
PR	04-JAN-2000; 2000US-174437P.		
XX			
XX	(ABEL/) ABEL L M.		
PA	(FALC/) FALCO S C.		
PA	(FAMO/) FAMODU O O.		
XX			
PI	Abell LM, Falco SC, Famodu OO;		
XX			
DR	WPI; 2002-113395/15.		
XX	P-PSDB; ABB08413.		
PT			
XX			
XX			
PS	Claim 7; Page 33; 57pp; English.		
XX			
CC	The invention relates to new isolated polypeptides, which comprises an		
CC	acetylactate synthase (ALS) small subunit. Acetylactate synthase is the		
CC	first committed step in branched chain amino acid biosynthesis in plants		


```

|||||
Db 252 CGGAGACACCCCTTACCATCGAGTACGTGCGGCAAAATTGCTGTGCA 311
QY 692 AAGAATTGAAAGATTCAGATCAGAGAGATTGTAAGACAGAAAGATACACTGAG 751
Db 312 AAGGAACCTCAGCAAAATTGGGAAAAAGAAATTGTANACGGAATAATGCTTTGA 371
QY 752 AAGGAAAAAGA 762
Db 372 GACGTGAAAAA 382

RESULT 14
ABA98821
ID ABA98821 standard; DNA; 417 BP.
XX
AC ABA98821;
XX
DT 18-JUN-2002 (first entry)
XX
DE Wheat ALS small subunit (clone wdk2c.pK015.a13) EST fragment.
XX
KM ALS: Acetolactate synthase; acetohydroxy synthase;
KM amino acid biosynthesis; plant; herbicide; ALS small subunit;
KM ALS holoenzyme; crop protection chemical; expressed sequence tag;
KM EST; ss.
XX
OS Triticum aestivum.
XX
FH Key Location/Qualifiers
FT CDS 1..408
FT /*tag= a
FT /partial
FT /product= "ALS small subunit"
FT /EC_number= "4.1.3.18"
FT /transl_except= (pos:265..267, aa:Xaa)
FT /transl_except= (pos:274..276, aa:Xaa)
FT /transl_except= (pos:343..345, aa:Xaa)
FT /transl_except= (pos:367..369, aa:Xaa)
FT /note= "Xaa is unknown; first 3 translation exceptions
FT encode in frame stop codons; no start codon present"
XX
PN US200104939-A1.
XX
PD 22-NOV-2001.
XX
PF 08-DEC-2000; 2000US-0732618.
XX
PR 04-JAN-2000; 2000US-174437P.
XX
PA (ABEL/) ABELL L M.
PA (FALCO/) FALCO S C.
PA (FAMODU/) FAMODU Q O.
XX
PI Abell LM, Falco SC, Famodu OO;
XX
DR MPI: 2002-113395/15.
DR P-PSDB; ABB08417.
XX
PT New small subunits of plant acetolactate synthase and nucleic acids
PT encoding them, useful in screening for novel crop protection chemicals
PT or potential herbicidal compounds based upon holoenzyme inhibition -
XX
PS Claim 7; Page 39; 57pp; English.
XX
CC The invention relates to new isolated polypeptides, which comprises an
CC acetylactate synthase (ALS) small subunit. Acetylactate synthase is the
CC first committed step in branched chain amino acid biosynthesis in plants
CC and bacteria. The polypeptides are useful in aiding the discovery of new
CC herbicides that inhibit plant acetylactate synthase activity. The
CC polypeptides are useful for preparing plant ALS holoenzyme, which are
CC useful in screening for potential herbicidal compounds based upon
CC holoenzyme inhibition. The ALS holoenzyme is also useful in screening

```

```

CC for novel crop protection chemicals. Particularly, the polypeptides can
CC be used as a target to facilitate design and/or identification of
CC inhibitors of those enzymes that may be useful as herbicides. Thus, the
CC polypeptides could be appropriate for new herbicide discovery and
CC design. The polynucleotides are also useful for producing the
CC polypeptides and for screening for new herbicides. The current sequence
CC represents the wheat ALS small subunit (clone wdk2c.pK015.a13) EST
CC fragment.
XX
SQ Sequence 417 BP; 127 A; 82 C; 107 G; 99 T; 2 other;
XX
Query Match 8.9%; Score 149.4; DB 24; Length 417;
Best Local Similarity 70.8%; Pred. No. 9,2e-36;
Matches 226; Conservative 0; Mismatches 91; Indels 2; Gaps 2;

QY 363 GCAGAGAGAGATACAAATATGAGAGTCTGCTGCTGCTGAGACAGACAAAGCTCA 422
Db 1 GCCAGGAGAGGCTACACATCGAGTCGTCGCCGTGGGCTTAACAGACAGAGGCCCTC 60
QY 423 TTCACCATAGTTGTCTGTGGAAGTGAAGGTAATTCACAGAGTCAATCGACAACTCAG 482
Db 61 TTCACCATGCTGCTCTCGGAGAGATAGGTCCTCAACAGACTCATTTAGCAGACTCAAT 120
QY 483 AAGCTGTTAATGTTCTAAAGTTGAAGATATCTCAAGTGAAGCCGCAAGTGAAGCGTGA 542
Db 121 AAGCTGTCACAGCTTGTGAATGTTGAAGATCTATCTAAGAGAGCTCAGGTTGAAGAGAG 180
QY 543 CTGATGCTGTTAAAGTGAATGCACATCCAGATCCAGGCGAGAGATCATGTGGCTAGTT 602
Db 181 CTCATCTTATTAACACTCAATGTTGAACCAAGTCAACAGTCGACGCTCATGTTTGAAGCT 240
QY 603 GACACATTC-AGAGCAAGAGTTGTAGATATAGGGAACATGCACTTGTACTATCGAGGTAC 661
Db 241 AATGTTTAAAGAGCAAAAGTTGTATATTTCTTGGAACAAGCTAACTGTGAGGTAAAC 300
QY 662 T-GGAGATCTCGGAAAAAT 679
Db 301 TGGAGAGATCTCGGAAAAAT 319

RESULT 15
AAC37251
ID AAC37251 standard; DNA; 420 BP.
XX
AC AAC37251;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 16718.
XX
KM Hybridisation assay; genetic mapping; gene expression control;
KM protein identification; signal transduction pathway;
KM metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.

```

PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151067.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155137.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157733.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158332.
PR 12-OCT-1999; 99US-0158836.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.

PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

	Query Match	8.8%	Score 147.6:	DB 21:	Length 420;
	Best Local Similarity	69.5%:	Pred. No. 3.3e-35;		
	Matches 198:	Conservative 1:	Mismatches 86;	Indels 0;	Gaps 0.
Oy	1234 TTTTCGAAAGACAACTGATAGCTGATTAGAATGGCCGTGAACCTCGTCTGTAGAGAGANG	1293			
Dd	8 TTTGGCGMGANAATGTGATGCTCATCAAAAGTAGCTGTGATACTTCTGCAGAGAGANG	67			
Oy	1294 TCCTTGACATTTGCTAGTATTTTTACAGGGCTAAAGCTGTGTACCCTATCTGATCACAATAA	1353			
Dd	68 TTTTAGATATTGCTCAAGTTTTCCTCGTGCTAAAGCCATTGATCTCTGATCACTACTATCA	127			
Oy	1354 CTTTTCACGTTACTGCGGGATCTAGACAAAGATGTTGCACACTGCCAAAAGTATTGAGCCCT	1413			
Dd	128 CCCTTGAGGTTACAGGGGATCTCAGAAAAATGTCTGCTACTCAAACGCAATTAGAGCCCT	187			
Oy	1414 ATGATATATGTAGSITTCGACAGAACCGGTCGTGTGGCATTTGGCTGTGATGGGAGTGG	1473			
Dd	188 ATGGAATCTGCAGAGTGCGTCCGACAGAGAGAGTGGCAATGGTAAGAGACATCTGAGAGTGG	247			
Oy	1474 ACTCGAAGTACTCGTGGATGACTCCTTTCTTTAAAGAGCTAAA	1518			
Dd	248 ATTGCACATTATCTGTGGATACCTCTCTTCCTTTAGTTCAAAA	292			

Search completed: March 9, 2003, 19:39:42
Job time : 444 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 09:00:40 ; Search time 40 Seconds
(without alignments)
1635.651 Million cell updates/sec

Title: US-09-997-900-2
Perfect score: 2444
Sequence: 1 MAASVSSSPSIRCLRSACS.....AREGVDSKYLRGVSFLLTG 491

Scoring table: BIOSIM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	length	DB ID	Description
1	2444	100.0	491	21	AA196203
2	2358.5	96.5	484	23	ABB91968
3	1692	69.2	476	23	ABB08416
4	1518	62.1	477	23	ABB3502
5	1493.5	61.1	483	23	ABB08412
6	1476	60.4	449	19	AAW79140
7	1395	57.1	365	23	ABB08415
8	1302	53.3	350	23	ABB08410
9	522	21.4	203	23	ABB08408
10	499	20.4	133	23	ABB08418

11	470.5	19.3	156	23	ABB08409
12	436	17.8	156	23	ABB08413
13	372.5	15.2	144	23	ABB08414
14	336	13.7	94	21	AA616178
15	332	13.6	86	21	AA616179
16	329.5	13.5	135	23	ABB08417
17	329.5	13.5	174	22	AAU01249
18	310.5	12.7	803	22	AA831992
19	307	12.6	172	22	AA693204
20	305	12.5	166	22	AAU37779
21	305	12.5	166	22	AAU38002
22	305	12.5	186	17	AAW06555
23	302.5	12.4	158	19	AAW85930
24	301	12.3	172	22	AA879360
25	277.5	11.4	163	22	AA667740
26	277	11.3	163	22	AAU36451
27	276.5	11.3	163	22	AAU34430
28	271.5	11.1	173	22	AAU38425
29	269.5	11.0	173	22	AAU36149
30	268.5	11.0	163	22	AAU35663
31	267.5	10.9	163	22	AAU33546
32	263.5	10.8	163	23	ABB48167
33	262.5	10.7	158	23	ABB54357
34	260.5	10.7	158	15	AA854221
35	138	5.6	43	23	ABB08411
36	137.5	5.6	104	22	AAU53173
37	135	5.5	828	22	AA691284
38	134.5	5.5	97	23	ABP07823
39	121	5.0	1138	14	AA837213
40	115	4.7	1583	22	AAU53954
41	112	4.6	1138	15	AA846225
42	111	4.5	495	22	AAU34722
43	110.5	4.5	737	21	AAW96614
44	110	4.5	507	22	AAW96433
45	107	4.4	500	22	AAU38171

ALIGNMENTS

RESULT 1	AA196203	standard; Protein; 491 AA.
ID	AA196203	
AC	AA196203	
DE	11-AUG-2000	(first entry)
XX	Arabidopsis acetoxy-acid synthase (AHAS) small subunit.	
XX	Herbicide-resistance; acetoxy-acid synthase; AHAS;	
XX	acetylacetate synthase; imidazolone; sulfonylurea;	
XX	triazolopyrimidine sulfonamide; sulfamoylurea; enzyme;	
XX	pyrimidyl-oxy-benzolic acid; sulfonylcarboxamide; transgenic plant;	
XX	branched-chain amino acid synthesis.	
OS	Arabidopsis sp.	
PN	WO200026390-A2.	
XX	11-MAY-2000.	
PD	28-OCT-1999;	99WO-US25452.
XX	29-OCT-1998;	98US-0106239.
PR	(AMCY) AMERICAN CYANAMID CO.	
PA	Kakefuda G, Costello C, Sun M, Hu W;	
XX	WPI: 2000-365633/31.	
DR	N-PSDB: AAA27424, AAA27425.	
XX		

Corn ALS small sub
Rice ALS small sub
Rice ALS small sub
Arabidopsis thaliana
Arabidopsis thaliana
Wheat ALS small su
B. subtilis acetoh
Amino acid sequenc
C glutamicum prote
Streptococcus pneu
Streptococcus pneu
R. capsulatus acet
S. pneumoniae deti
Corynebacterium q1
Small subunit of i
pseudomonas aerugi
E. coli cellular p
Salmonella typhi c
Klebsiella pneumon
Haemophilus influe
Klebsiella pneumon
Listeria monocytog
Lactococcus lactis
L.lactis alpha-ace
Corn ALS small sub
Propionibacterium
C glutamicum prote
Human ORFX protein
B.t. toxin HD511.
Propionibacterium
Bacillus thuringie
E. coli cellular p
Streptococcus pneu
putative P. abyssal
Salmonella typhi c

PT New polynucleotide encoding eukaryotic acetoxyhydroxy-acid synthetase
PT small subunit protein for producing transgenic herbicide resistant
PT plants and identifying mutations affecting enzymatic activity of the
PT synthetase
PS Disclosure: Page 50-52; 57pp; English.
XX
XX
CC The present sequence is the Arabidopsis acetoxyhydroxy-acid synthase (AHAS)
CC small subunit protein. AHAS is also known as acetolactate synthase. This
CC enzyme is needed for branched-chain amino acid synthesis and so is
CC essential for life. Inhibition of this enzyme would lead to plant death
CC and therefore inhibitors would be potential herbicides. Certain
CC herbicides are known to inhibit AHAS: imidazolinones, sulfonlureas,
CC triazopyrimidine sulfonamides, pyrimidyl-oxy-benzoic acids,
CC sulfamoylureas and sulfonycarboxamides. The present sequence may be used
CC to identify mutant AHAS which are resistant to these herbicides and
CC may be used to create herbicide resistant transgenic plants e.g. dicot
CC and monocot crop plants.
XX
XX
SQ Sequence 491 AA:

Query Match 100.0%; Score 2444; DB 21; Length 491;
Best Local Similarity 100.0%; Pred. No. 2e-220;
Matches 491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX
OS Arabidopsis thaliana.
XX
XX
PN WO200210210-A2.
XX
PD 07-FEB-2002.
XX
PF 28-AUG-2001; 2001WO-EP09892.
XX
PR 28-AUG-2001; 2001WO-EP09892.
XX
PA (FARB) BAYER AG.
XX
PI Tietjen K, Weidler M;
DR WPI: 2002-269010/31.
XX
PT Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -
PS Claim 5; SEQ ID NO 1179; 261pp + Sequence Listing; English.
XX
XX
CC The invention relates to identifying target proteins
CC (AB990790-AB94016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.
XX
SQ Sequence 484 AA:

Query Match 96.5%; Score 2358.5; DB 23; Length 484;
Best Local Similarity 97.2%; Pred. No. 2.1e-212;
Matches 480; Conservative 0; Mismatches 1; Indels 13; Gaps 2;

Best Local Similarity 69.8%; Pred. No. I.5e-133;
Matches 296; Conservative 59; Mismatches 65; Indels 4; Gaps 2;

```

QY 68 SASPSEASAPPKS--KVRKHTISYFVGDESGMINRIGVAFARRGYNIESLAVGINRDK 125
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 56 TKKAITTAOSVAP7ACDVRKRHTISYFVGDESGINRIGVAFARRGYNIESLAVGINRDK 115
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 126 ALFTTIVCGTERVLOQVIEPOLKLVNLKVEDISSEPOVERELMLVKVAHPSRAEIMM 185
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 116 ALFTTIVGCTKVLQOVVQNLKLVNLKVEDLSKPHRELMILKLNADSTREIMM 175
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 186 LVDTFARAVYDIAEHALTEVFGDPCKMIAVERNLKKFOIREIVTGKTALRREKMGATA 245
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 176 LVDFIRAKIVDTSSEOSTLEVTEGDPCKMVALTTNLEKFGIKETIARFKTALRREKMGATA 235
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 246 PFMRESAASYPDLKEGAPSVLRSSKKGAIVPOKETSAGGDVYVPEPFDPVYHRLDHA 305
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 236 PFMRESAASYPDLKEGAPSVLRSSKKGAIVPOKETSAGGDVYVPEPFDPVYHRLDHA 293
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 306 WGLLTDDETSGLRSHHTLSLVNDIPGVNLIVTVGFARRGYNIOSLAVGHAETKGISRTT 365
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 294 WGMVVDDESSGLRSHHTLSLVNAPGVNLITGALSRRGYNIOSLAVGHAETKGISRTT 353
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 366 VIPATDESYSKLVQQLYKLVNVEVHDLTHLPFSERELMLKIAVNAARRDVLDIASIF 425
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 354 VIPGTGENDIKLVROKLIDLOEIONITHMPFAERELMLKIAVADTSARRDVLDIAQVF 413
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 426 RAKAVDVSHTTTLQTLTGDLCKMVALQRLLEPYGICEVARTGRVALARSSGVDSKYLRG 485
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 414 RAKAIDVSDHTTLEVTGDLRKMVALQTLQLEAVGICEVARTGRVALARSSGVDSKYLRG 473
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 486 SFLL 489
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 474 SLPL 477
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 5
ABB08412
ID ABB08412 standard; protein; 483 AA.
XX
AC ABB08412;
XX
DT 18-JUN-2002 (first entry)
XX
DE Corn ALS small subunit (clone p0094.cssl172ra).
XX
KW ALS: Acetolactate synthase; EC 4.1.3.18; acetohydroxy synthase;
   amino acid biosynthesis; plant; herbicide; ALS small subunit;
   ALS holoenzyme; crop protection chemical; enzyme.
XX
OS Zea mays.
XX
PN US2001044939-A1.
XX
PD 22-NOV-2001.
XX
PE 08-DEC-2000; 2000US-0732618.
XX
PR 04-JAN-2000; 2000US-174437P.
XX
PA (ABELL) ABELL L M.
   (FALC) FALC S C.
   (FAMO) FAMODU O O.
XX
PI Abell LM, Falco SC, Famodu OO;
XX
DR WPI; 2002-113395/15.
XX
DT N-PSDB; ABA98816.
XX
DE New small subunits of plant acetolactate synthase and nucleic acids
   encoding them, useful in screening for novel crop protection chemicals
   or potential herbicidal compounds based upon holoenzyme inhibition
XX

```

PS Claim 19; Fig 3; 57pp; English.

XX The invention relates to new isolated polypeptides, which comprises an
 CC acetolactate synthase (ALS) small subunit. Acetolactate synthase is the
 CC first committed step in branched chain amino acid biosynthesis in plants
 CC and bacteria. The polypeptides are useful in aiding the discovery of new
 CC herbicides that inhibit plant acetolactate synthase activity. The
 CC polypeptides are useful for preparing plant ALS holoenzyme, which are
 CC useful in screening for potential herbicidal compounds based upon
 CC holoenzyme inhibition. The ALS holoenzyme is also useful in screening
 CC for novel crop protection chemicals. Particularly, the polypeptides can
 CC be used as a target to facilitate design and/or identification of
 CC inhibitors of those enzymes that may be useful as herbicides. Thus, the
 CC polypeptides could be appropriate for new herbicide discovery and
 CC design. The polynucleotides are also useful for producing the
 CC polypeptides and for screening for new herbicides. The current sequence
 CC represents the corn ALS small subunit (clone p0094.cssl172ra).

SQ Sequence 483 AA;

Query Match 61.1%; Score 1493.5; DB 23; Length 483;

Best Local Similarity 63.5%; Pred. No. 3.1e-131;
Matches 308; Conservative 61; Mismatches 83; Indels 33; Gaps 7;

```

QY 13 RCLRSACSDSSPALVSTVSPSPAKISYLSGISHRGDEMGRMGFVSVDKISDASF 72
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 24 RCGAGSRVDSRPL---TPAVGFTA-----GPR---ARSV--AVTAASS 58
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 73 SEAS-SATP-----KSKVRKHTISYFVGDESGMINRIGVAFARRGYNIESLAVGINRDK 125
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 59 SPATGVPFVPPRSNRSYVAKRHHTLSYFVGDESGMINRIGVAFARRGYNIESLAVGINRDK 118
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 126 ALFTTIVCGTERVLOQVIEPOLKLVNLKVEDISSEPOVERELMLVKVAHPSRAEIMM 185
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 119 ALFTTIVGCTKVLQOVVQNLKLVNLKVEDLSKPHRELMILKLNADSTREIMM 178
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 186 LVDTFARAVYDIAEHALTEVFGDPCKMIAVERNLKKFOIREIVTGKTALRREKMGATA 245
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 179 LVIRIFAEVVDSDYTLTTEVTGDPCKMVALQTLKSGYIARIETGKTALRREKMGATA 238
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 246 PFMRESAASYPDLKEGAPSVLRSSKKGAIVPOKETSAGGDVYVPEPFDPVYHRLDHA 304
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 239 PFMRESAASYPDLKEVAPIFNQONTGVKAINONPESSSGDVYVPESSYFSSSTQILDA 298
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 305 HNGLLTDETSGLRSHHTLSLVNDIPGVNLIVTVGFARRGYNIOSLAVGHAETKGISRTT 364
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 299 HNGVMTDGPPTGFCSTHTLSLVNDIPGVNLIVTVGFARRGYNIOSLAVGHAETKGISRTT 358
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 365 TVIPATDESYSKLVQQLYKLVNVEVHDLTHLPFSERELMLKIAVNAARRDVLDIASI 424
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 359 TVVPGTDESIAKLVLQYKLVNVEVQDFTHLPPFAERELMLKIAVNAARRDVLDIAQI 418
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 425 FRAKAVDVSHTTTLQTLTGDLCKMVALQRLLEPYGICEVARTGRVALARSSGVDSKYLRG 484
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 419 FEAQKVDISDHTTLLTLDIDRMVRLQMLEQYGICEVARTGRVALARSSGVDSKYLRG 478
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 485 YSFL 489
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 479 FSUPL 483
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 6
AAW79140
ID AAW79140 standard; protein; 449 AA.
XX
AC AAW79140;
XX
DT 18-NOV-1998 (first entry)
XX
DE Plant acetolactate synthase (ALS) small subunit protein.
XX
KW ALS: small subunit; acetolactate synthase; plant; ssuALS; 1suALS;
   Nicotiana glumbaginifolia; plasmid vector; herbicide; holoenzyme;
XX

```


KW large subunit.
 XX
 OS ~~Nicotiana glauca~~ *Nicotiana glauca*.
 XX
 PN MO9837206-A1.
 XX
 PD 27-AUG-1998.
 XX
 PF 23-FEB-1998; 98MO-US03506.
 XX
 PR 24-FEB-1997; 97US-0039148.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E. I.
 XX
 PI Abell LM, Hershey HP;
 XX
 DR WPI: 1998-467568/40.
 XX
 N-PSDB; AAV53857.
 XX
 PR New nucleic acid encoding a plant aceto:lactate synthase small
 PT sub:unit - that combines with the large sub:unit to give a
 PT holoenzyme having higher activity than the large sub:unit alone
 XX
 PS Claim 1: Pages 30-32; 47pp; English.

CC This represents a small subunit of a plant acetylactate synthase
 CC (ssuALS). The encoding cDNA is contained in the plasmid PSSU.NPI. The
 CC plasmid vector comprising the ssuALS nucleotide operably linked to a
 CC regulatory sequence can be used to transform host cells for the
 CC recombinant production of the protein which is used for evaluating a
 CC compound for acetylactate synthase inhibition and so for selecting
 CC potential herbicides. Evaluation of a compound for acetylactate synthase
 CC (ALS) inhibition comprises expression and purification of plant ssuALS
 CC from the transformed host, mixing this ssuALS with the large subunit of
 CC ALS to form a holoenzyme which is then treated with a test compound.
 CC Treated and untreated holoenzyme activity are compared to select
 CC compounds with potential for herbicidal activity. Previously, large
 CC subunits of plant ALS (lsuALS) have been isolated, but the existence of
 CC ssuALS had not been verified. Mixing the two subunits results in a
 CC holoenzyme that has a 4-15 fold increase in specific activity over
 CC lsuALS alone.

XX Sequence 449 AA:

Query Match 60.4%; Score 1476; DB 19; Length 449;

Best Local Similarity 79.1%; Pred. No. 1,2e-128; Mismatches 42; Indels 0; Gaps 0;

Matches 291; Conservative 85; Mismatches 42; Indels 0; Gaps 0;

OY 99 MINRIAGVFARGNYTESLAVGLNDRKALFTIVCGTETVLOQVTEOLQKLVNVLKVEDI 158
 DB |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 37
 37 MINRIAGVFARGNYTESLAVGLNDRKALFTIVCGTETVLOQVTEOLQKLVNVLKVEDI 96
 OY 159 SEPOVERELMLVKYNAHESRAELMVLDTFRARVVDIAEHALITEVTDGDEKMAVER 218
 DB |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 97
 97 SEPOVERELMLVKYNAHESRAELMVLDTFRARVVDIAEHALITEVTDGDEKMAVER 156
 OY 219 NKKFOIRIVTFRGKIALREKMGATAPWRFSAASTPYLKGQAPVSVLRSSKKGAVPQ 278
 DB |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 157
 157 NKKFOIRIVTFRGKIALREKMGATAPWRFSAASTPYLKGQAPVSVLRSSKKGAVPQ 216
 OY 279 KETSAAGDVPVEPFDPKVRHILDAHMGILTDEDTSGLRSHSLVNDIGVNLIVTG 338
 DB |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 217
 217 KETSAAGDVPVEPFDPKVRHILDAHMGILTDEDTSGLRSHSLVNDIGVNLIVTG 276
 OY 339 VVARRGYNQSLAVGHAETKGSRTITVTPATDEYSKILVQOLKLVADVHEVHDLTHLDF 398
 DB |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 277
 277 VVARRGYNQSLAVGHAETKGSRTITVTPATDEYSKILVQOLKLVADVHEVHDLTHLDF 336
 OY 399 SERELMLIKIYNAARVDLIASIFRAKAVDSOHTITLDTGLDKMVALORLEPEY 458
 DB |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 337
 337 SERELMLIKIYNAARVDLIASIFRAKAVDSOHTITLDTGLDKMVALORLEPEY 396
 OY 459 GICEVART 466

DB 397 GICEVART 404

RESULT 7
 ABB08415
 ID ABB08415 standard; protein; 365 AA.

AC ABB08415;
 DT 18-JUN-2002 (first entry)

DE Rice ALS small subunit (clone r10npk117.a16).

KW ALS: Acetylactate synthase; EC 4.1.3.18; acetylhydroxy synthase;
 KM amino acid biosynthesis; plant; herbicide; ALS small subunit;
 KM ALS holoenzyme; crop protection chemical; enzyme.

OS Oryza sativa.

PN US2001044939-A1.

PD 22-NOV-2001.

PF 08-DEC-2000; 2000US-0732618.

PR 04-JAN-2000; 2000US-174437P.

PA (ABELL) ABELL L M.
 PA (FALCO) FALCO S C.
 PA (FAMO) FAMODU O O.

PI Abell LM, Falco SC, Famodu OO;

DR WPI: 2002-113395/15.
 N-PSDB; ABA98819.

PT New small subunits of plant acetylactate synthase and nucleic acids
 PT encoding them, useful in screening for novel crop protection chemicals
 PT or potential herbicidal compounds based upon holoenzyme inhibition

PS Claim 19; Page 36; 57pp; English.

CC The invention relates to new isolated polypeptides, which comprises an
 CC acetylactate synthase (ALS) small subunit. Acetylactate synthase is the
 CC first committed step in branched chain amino acid biosynthesis in plants
 CC and bacteria. The polypeptides are useful in aiding the discovery of new
 CC herbicides that inhibit plant acetylactate synthase activity. The
 CC polypeptides are useful for preparing plant ALS holoenzyme, which are
 CC useful in screening for potential herbicidal compounds based upon
 CC holoenzyme inhibition. The ALS holoenzyme is also useful in screening
 CC for novel crop protection chemicals. Particularly, the polypeptides can
 CC be used as a target to facilitate design and/or identification of
 CC inhibitors of those enzymes that may be useful as herbicides. Thus, the
 CC polypeptides could be appropriate for new herbicide discovery and
 CC design. The polynucleotides are also useful for producing the
 CC polypeptides and for screening for new herbicides. The current sequence
 CC represents the rice ALS small subunit (clone r10n.pk117.a16).

XX Sequence 365 AA:

Query Match 57.1%; Score 1395; DB 23; Length 365;

Best Local Similarity 75.6%; Pred. No. 3.6e-122; Mismatches 49; Indels 2; Gaps 1;

Matches 276; Conservative 38; Mismatches 49; Indels 2; Gaps 1;

OY 127 LFTIVCGTETVLOQVTEOLQKLVNVLKVEDISSPOVERELMLVKYNAHESRAELMVL 186
 DB |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 1
 1 LFTIVCGTETVLOQVTEOLQKLVNVLKVEDISSPOVERELMLVKYNAHESRAELMVL 60
 OY 187 VDTFRARVVDIAEHALITEVTDGDEKMAVERNLKRFQIREIVTFRGKIALREKMGATAP 246
 DB |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 61
 61 VDTFRARVVDIAEHALITEVTDGDEKMAVERNLKRFQIREIVTFRGKIALREKMGATAP 120


```

XX Abell LM, Falco SC, Famodu OO;
PI
XX WPI: 2002-113395/15.
DR
XX New small subunits of plant acetolactate synthase and nucleic acids
PT encoding them, useful in screening for novel crop protection chemicals
PT or potential herbicidal compounds based upon holoenzyme inhibition
XX
PS Disclosure; Page 25-26; 57pp; English.
XX
CC The invention relates to new isolated polypeptides, which comprises an
CC acetolactate synthase (ALS) small subunit. Acetolactate synthase is the
CC first committed step in branched chain amino acid biosynthesis in plants
CC and bacteria. The polypeptides are useful in aiding the discovery of new
CC herbicides that inhibit plant acetolactate synthase activity. The
CC polypeptides are useful for preparing plant ALS holoenzyme, which are
CC useful in screening for potential herbicidal compounds based upon
CC holoenzyme inhibition. The ALS holoenzyme is also useful in screening
CC for novel crop protection chemicals. Particularly, the polypeptides can
CC be used as a target to facilitate design and/or identification of
CC inhibitors of those enzymes that may be useful as herbicides. Thus, the
CC polypeptides could be appropriate for new herbicide discovery and
CC design. The polynucleotides are also useful for producing the
CC polypeptides and for screening for new herbicides. The current sequence
CC represents the corn ALS small subunit (clone m15.12.b12.sk20).
CC
SQ Sequence 203 AA;

```

Query Match 21.4%; Score 522; DB 23; Length 203;
 Best Local Similarity 60.1%; Pred. No. 1.3e-40;
 Matches 107; Conservative 27; Mismatches 40; Indels 4; Gaps 2;

```

OY 281 TSAGDVPVEFPFVKRIIDAHMGLTDEDTSGLRSHTSLIVNDIPGVLNTGVF 340
    ||||||| : : : ||||| : : : ||||||| : : : |||||||
DB 24 SNAGDVPVEYESLSVNHVLDAMHGVLDGDDATGLRSHTSLIVNDIPGVLNTGVF 83
OY 341 ARGYNISLAVGNAETKGISRTITVPATDESVSRLVQOLKLVDFEVDL---THLP 397
    ||||||| : : : ||||||| : : : ||||||| : : : |||||||
DB 84 ARGYNISLAVGPAKEKESIRITVPPTVESIEKVSASTSLMCMESSMTLEPHL- 142
OY 398 FSEBELMIKIVNNAARDVLDIASIFRAKAVDYSDHTITLQTLGDDKVAOLRL 455
    : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 143 LKGNLSISKVSVMNAARKEILHIAQIFRAKPDVSDHTVTLTGDDDKMVDTKVI 200

```

RESULT 10
 ABB08418
 ID ABB08418 standard; Protein; 133 AA.
 AC ABB08418;
 DT 18-JUN-2002 (first entry)
 DE Wheat ALS small subunit (clone wdk2c.pk015.a13).
 XX ALS: Acetolactate synthase; acetoxyhydroxy synthase;
 KM amino acid biosynthesis; plant; herbicide; ALS small subunit;
 KW ALS holoenzyme; crop protection chemical; enzyme.
 XX Triticum aestivum.
 OS US2001044939-A1.
 PN
 XX US2001044939-A1.
 PD 22-NOV-2001.
 XX 08-DEC-2000; 2000US-0-32618.
 PF 04-JAN-2000; 2000US-174437P.
 PR (ABELL) ABELL L M.
 PA (FALCO) FALCO S C.
 PA (FAMO) FAMODU O O.

```

XX Abell LM, Falco SC, Famodu OO;
PI
XX WPI: 2002-113395/15.
DR N-PSDB: ABA98822.
XX
XX New small subunits of plant acetolactate synthase and nucleic acids
PT encoding them, useful in screening for novel crop protection chemicals
PT or potential herbicidal compounds based upon holoenzyme inhibition
XX
PS Claim 19; Page 40; 57pp; English.
XX
CC The invention relates to new isolated polypeptides, which comprises an
CC acetolactate synthase (ALS) small subunit. Acetolactate synthase is the
CC first committed step in branched chain amino acid biosynthesis in plants
CC and bacteria. The polypeptides are useful in aiding the discovery of new
CC herbicides that inhibit plant acetolactate synthase activity. The
CC polypeptides are useful for preparing plant ALS holoenzyme, which are
CC useful in screening for potential herbicidal compounds based upon
CC holoenzyme inhibition. The ALS holoenzyme is also useful in screening
CC for novel crop protection chemicals. Particularly, the polypeptides can
CC be used as a target to facilitate design and/or identification of
CC inhibitors of those enzymes that may be useful as herbicides. Thus, the
CC polypeptides could be appropriate for new herbicide discovery and
CC design. The polynucleotides are also useful for producing the
CC polypeptides and for screening for new herbicides. The current sequence
CC represents the wheat ALS small subunit (clone wdk2c.pk015.a13).
CC
SQ Sequence 133 AA;

```

Query Match 20.4%; Score 499; DB 23; Length 133;
 Best Local Similarity 73.3%; Pred. No. 9.6e-39;
 Matches 96; Conservative 22; Mismatches 13; Indels 0; Gaps 0;

```

OY 108 ARGYNIESLAVGLNRKALFTIYVCGTERTVLOQYIELOKLVNLRKVEDISSEPOVERE 167
    ||||||| : : : ||||||| : : : ||||||| : : : |||||||
DB 3 ARGYNIESLAVGLNRKALFTIYVCGTERTVLOQYIELOKLVNLRKVEDISSEPOVERE 62
OY 168 LMLVYNMHPESRAEIMLVDFEFARVVDIAERHALTIEVTPGPKMIAVERLKKFOIRE 227
    ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 63 LMLKLVNPEPDRAIDMFANVFRKAVVDISENSLTLEYTGPDKIVAAORLKRFGILE 122
OY 228 IVRTGKIALRR 238
    | ||||| :
DB 123 ICRGKIALRQ 133

```

RESULT 11
 ABB08409
 ID ABB08409 standard; Protein; 156 AA.
 AC ABB08409;
 DT 18-JUN-2002 (first entry)
 DE Corn ALS small subunit (clone cen3n.pk0112.c11 - EST fragment).
 XX ALS: Acetolactate synthase; EC 4.1.3.18; acetoxyhydroxy synthase;
 KM amino acid biosynthesis; plant; herbicide; ALS small subunit;
 KW ALS holoenzyme; crop protection chemical; enzyme.
 XX Zea mays.
 OS
 XX
 PH Key Location/Qualifiers
 FT Misc-difference 35 //Label= unknown
 FT //note= "Encoded by GNN"
 FT Misc-difference 43 //Label= unknown
 FT //note= "Encoded by GNC"
 FT Misc-difference 54 //note= "Encoded by GNN"
 FT Misc-difference 55

FT /label= unknown
 FT /note= "Encoded by GNT"
 FT Misc-difference 57
 FT /label= unknown
 FT /note= "Encoded by ANA"
 FT Misc-difference 58
 FT /label= unknown
 FT /note= "Encoded by NCT"
 FT Misc-difference 64
 FT /label= unknown
 FT /note= "Encoded by ANA"
 FT Misc-difference 66
 FT /label= unknown
 FT /note= "Encoded by GNA"
 FT Misc-difference 95
 FT /note= "Encoded by GTN"
 FT Misc-difference 98
 FT /note= "Encoded by GTN"
 FT Misc-difference 102
 FT /note= "Encoded by GTN"
 FT Misc-difference 119
 FT /label= unknown
 FT /note= "Encoded by ANA"
 FT Misc-difference 130
 FT /label= unknown
 FT /note= "Encoded by GNA"
 FT Misc-difference 149
 FT /label= unknown
 FT /note= "Encoded by NAT"
 FT Misc-difference 150
 FT /note= "Encoded by CCN"
 FT Misc-difference 151
 FT /label= unknown
 FT /note= "Encoded by GNG"
 PN US200104939-A1.
 XX
 XX 22-NOV-2001.
 PD
 XX 08-DEC-2000; 2000US-0732618.
 PF
 XX 04-JAN-2000; 2000US-174437P.
 PR
 XX (ABELL) ABELL L M.
 PA (FALCO) FALCO S C.
 PA (FAMO/) FAMODU O O.
 XX
 PI Abell LM, Falco SC, Famodu OO;
 XX
 DR WPI: 2002-113395/15.
 DR N-PSDB: ABA98813.
 XX
 PT New small subunits of plant acetylactate synthase and nucleic acids
 PT encoding them, useful in screening for novel crop protection chemicals
 PT or potential herbicidal compounds based upon holoenzyme inhibition -
 XX
 PS Claim 19; Page 27-28; 57pp; English.
 CC
 CC The invention relates to new isolated polypeptides, which comprises an
 CC acetylactate synthase (ALS) small subunit. Acetylactate synthase is the
 CC first committed step in branched chain amino acid biosynthesis in plants
 CC and bacteria. The polypeptides are useful in aiding the discovery of new
 CC herbicides that inhibit plant acetylactate synthase activity. The
 CC polypeptides are useful for preparing plant ALS holoenzyme, which are
 CC useful in screening for potential herbicidal compounds based upon
 CC holoenzyme inhibition. The ALS holoenzyme is also useful in screening
 CC for novel crop protection chemicals. Particularly, the polypeptides can
 CC be used as a target to facilitate design and/or identification of
 CC inhibitors of these enzymes that may be useful as herbicides. Thus, the
 CC polypeptides could be appropriate for new herbicide discovery and
 CC design. The polynucleotides are also useful for producing the
 CC polypeptides and for screening for new herbicides. The current sequence
 CC represents the corn ALS small subunit (clone cen3n.pk0112.c11) encoded

CC by an EST fragment.
 XX
 SQ Sequence 156 AA:
 Query Match 19.3%; Score 470.5; DB 23; Length 156;
 Best Local Similarity 63.5%; Pred. No. 5.7e-36;
 Matches 99; Conservative 15; Mismatches 37; Indels 5; Gaps 2;
 QY 141 QVIEOLKLVNLYKVEDISEPOVERELMLVKNAPESRAEIMLVDFRFRVYDIAEH 200
 DB 1 QVIEOLNKLNVNHSVEDLSKEPOVERELMLITLKNXPPDREXMYLVDFRKYVDXEXK 60
 QY 201 ALTIEVTGDPGKMIANVERNLKKFQIREIVRTGKIALRREKMGATAPFWFSAASYPDLKE 260
 DB 61 TLTXEAGDPGKIAAVQRLRKFGIKEICRTGKIALRREKIGATAFRWRFSASYPDLXE 120
 QY 261 QAPVSYLRSKSK---GAIYPQKETSAGGVYVEPE 293
 DB 121 ALPKKPLTSMKNTVNGSFV--RPSNAGDVPXPESY 154
 RESULT 12
 ABB08413
 ID ABB08413 standard; protein; 156 AA.
 XX
 AC ABB08413;
 XX
 DT 18-JUN-2002 (first entry)
 XX
 DE Rice ALS small subunit (clone r10.pk084.a24 - EST fragment).
 XX
 KW ALS: Acetylactate synthase; EC 4.1.3.18; acetylhydroxy synthase;
 KW amino acid biosynthesis; plant; herbicide; ALS small subunit.
 KW ALS holoenzyme; crop protection chemical; enzyme.
 XX
 OS Oryza sativa.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 154
 FT /note= "Encoded by CCN"
 FT
 PN US200104939-A1.
 XX
 XX 22-NOV-2001.
 PD
 XX 08-DEC-2000; 2000US-0732618.
 PF
 XX 04-JAN-2000; 2000US-174437P.
 PR
 XX (ABELL) ABELL L M.
 PA (FALCO) FALCO S C.
 PA (FAMO/) FAMODU O O.
 XX
 PI Abell LM, Falco SC, Famodu OO;
 XX
 DR WPI: 2002-113395/15.
 DR N-PSDB: ABA98817.
 XX
 PT New small subunits of plant acetylactate synthase and nucleic acids
 PT encoding them, useful in screening for novel crop protection chemicals
 PT or potential herbicidal compounds based upon holoenzyme inhibition -
 XX
 PS Claim 19; Page 33-34; 57pp; English.
 CC
 CC The invention relates to new isolated polypeptides, which comprises an
 CC acetylactate synthase (ALS) small subunit. Acetylactate synthase is the
 CC first committed step in branched chain amino acid biosynthesis in plants
 CC and bacteria. The polypeptides are useful in aiding the discovery of new
 CC herbicides that inhibit plant acetylactate synthase activity. The
 CC polypeptides are useful for preparing plant ALS holoenzyme, which are
 CC useful in screening for potential herbicidal compounds based upon
 CC holoenzyme inhibition. The ALS holoenzyme is also useful in screening
 CC for novel crop protection chemicals. Particularly, the polypeptides can

CC be used as a target to facilitate design and/or identification of
CC inhibitors of those enzymes that may be useful as herbicides. Thus, the
CC polypeptides could be appropriate for new herbicide discovery and
CC design. The polynucleotides are also useful for producing the
CC polypeptides and for screening for new herbicides. The current sequence
CC represents the rice ALS small subunit (clone fl0n.pk084.a24) encoded
CC by an EST fragment.

SQ Sequence 156 AA;

Query Match	17.8%;	Score 436;	DB 23;	Length 156;
Best Local Similarity	60.1%;	Pred. No. 9.9e-33;		
Matches	86;	Conservative	23;	Mismatches 32;
			Indels	2;
			Gaps	1

QY	173	VNAHESRAELMLVDFERRARVVDIAHALHLEVTGPGKMIIVERLULKEFOJREIYRTG	232
QY	1	LTVPPDORPEVMVLVDLFRAKVVDSENTLTIIEVTGPGKIVAAQRLSLFSGKEICRTG	60
Db	233	KIALRREMGATAPFMFSSAASYDLKEQAPSVLSRSGKCAIYP--QKETSAGGVYYP	260
QY	61	KIALRREKIGATAPFMFSSAASYDLLEALPKNSLTSVMKTYNGSTDOSSNAGGVYYP	120
Db	291	EPFFDPKVRHLIDAHKGLDDED	313
QY	121	EPYCGSSNOVLDAHMGVLDED	143
Db			

RESULT 13

ID ABB08414 standard; protein; 144 AA.

AC ABB08A14;
XX 18-JUN-2002 (first entry)
DT
XX
DE Rice ALS small subunit (Stone 10npk117.a16 - EST fragment).
XX
KW ALS: Acetylactate synthase; EC 4.1.3.18; acetylhydroxy synthase;
KW amino acid biosynthesis; plant; herbicide; ALS small subunit;
KW ALS holoenzyme; crop protection chemical; enzyme.

OS *Oryza sativa*.

FH	Key	Location/Qualifiers
FM	3:55pm	114

	/label= unknown
E ₁ I ₁	/note= "Encoded by NNA"
ET	

F1	Misc-allele IS/
FT	/label] = unknown

11	FT	Misc-difference	138	7/10/68	Encoded by me
----	----	-----------------	-----	---------	---------------

FT /note= "Encoded by TGN"

FT	/label = unknown
FT	

FT	Misc-difference	144	/1abc] = unknown
FT			

```

/Note= "Encoded by NCA"
FI
XX

```

PN 052001044939-AL-
XX

FD 22-NOV-2001
XX

XX

XX

PA	(FALC/)	FALCO S C.
PA	(FAMO/)	FAMODU O O

PI Abell LM, Falco SC, Famodu OO;

XX	WPI; 2002-113395/15.
DR	
DR	N-PSDB; ABA98818.
DR	

PT New small subunits of plant acetylcoA synthase and nucleic acids
PT encoding them, useful in screening for novel crop protection chemicals
PT or potential herbicidal compounds based upon holoenzyme inhibition -

PS Claim 19; Page 34-35; 57pp; English.

The invention relates to new isolated polypeptides, which comprises an acetoactate synthase (ALS) small subunit. Acetoactate synthase is the first committed step in branched chain amino acid biosynthesis in plants and bacteria. The polypeptides are useful in aiding the discovery of new herbicides that inhibit plant acetoactate synthase activity. The polypeptides are useful for preparing plant ALS holoenzyme, which are useful in screening for potential herbicidal compounds based upon holoenzyme inhibition. The ALS holoenzyme is also useful in screening for novel crop protection chemicals. Particularly, the polypeptides can be used as a target to facilitate design and/or identification of inhibitors of those enzymes that may be useful as herbicides. Thus, the polypeptides could be appropriate for new herbicide discovery and design. The polynucleotides are also useful for producing the polypeptides and for screening for new herbicides. The current sequence represents the rice ALS small subunit (clone p10n.pK117.a16) encoded by an EST fragment.

SQ	Sequence	1.44	AA;
----	----------	------	-----

Query Match	15.2%	Score 372.5;	DB 23;	Length 144;
Best Local Similarity	60.6%	Pred. No. 8e-27;		
Matches	86;	Conservative 15;	Mismatches 24;	Indels 17;
				Gaps 5

```
QY      119 VGLNDKALFTIVVCGIERVLQOYIEQLKLNVNLVKVEDISSBPQVERLMLVKVNAHPE 17
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:
Db       2 VGLNDKAMFTIVSGTDRLNVOYEQLKNLVNLVNLEDLSKEPQVEEDLMLIKINVEPD 61
```

```

QY 179 SRAELMWLVTFRARVVDIAEALTIETGDP-GKMI---AVERNL--KKQIREIVTGT 232
      |::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 62 QRPEVMVLVDIFRAKVVDISENTLTIEYGRSAKLLLCGTSAIDGKKKFXRE----- 116

```

```
QY      233 KIALREKMGATAPWRFSAAS 254
          |:|||||: | |||
Db      117 KIALREKI-----WEQLAAS 132
```

RESULT 14

ID	AA
AAG16178	standard; Protein; 94 AA.

AC AAG16178;
XY .

DI 17-001-2000 (first entry)
XX

22 HUMANITARIAN PROTECTIVE DIAGNOSTIC SEX ID NO: 10/13
XX

KW protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

US Arablaopsis enallana
XX

XX
EF

AZ
COCTE

XX

XX

0
7
J
E
9
A
O
O
O

9
9
I
I

2
A
O
A
O
O
O

PR	05-MAR-1999;	99US-0123180
PR	06-MAR-1999;	99US-0123180

PR	23-MAR-1999;	99US-0125788
PR	25-MAR-1999;	99US-0126264

PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130049.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148341.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149829.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.

PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0153633.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.

PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 13.6%; Score 332; DB 21; Length 86;
Best Local Similarity 77.9%; Pred.No. 2,3e-23;
Matches 67; Conservative 7; Mismatches 12; Indels 0; Gaps 0;
QY 404 MLKIAVNAARDVDLDIASIFRAKAVDSHTITLQLTGDDKMWALORLEPYGICEV 463
DB 1 MLKVAADTSARDVDLDIQVRAKALDVSHTITLEVYGDLRKKSALOTQLEATYGICEV 60
QY 464 ARTGRVALARESGVDSKYRGYSFLL 489
DB 61 ARTGRVALVRESGVDSYLRGYSPL 86

Search completed: March 10, 2003, 09:07:58
Job time : 43 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 09:06:30 ; Search time 21 Seconds

(without alignments)
2247.714 Million cell updates/sec

Title: US-09-997-900-2

Perfect score: 2444

Sequence: 1 MAATSVSSSPSIRCLRSACS.....ARESGVDSKYLGRYSFLITG 491

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: pIR_73:*
2: pIR_1:*
3: pIR_2:*
4: pIR_4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2358.5	96.5	484	2	D84725
2	463	18.9	182	2	AC3384
3	461	18.9	188	2	S74358
4	427	17.5	174	2	S73151
5	359.5	14.7	159	2	F69464
6	359.5	14.7	192	2	E70459
7	351	14.4	168	2	E70855
8	345	14.1	174	2	T35829
9	341	14.0	169	2	T45414
10	329.5	13.5	174	2	E69644
11	327.5	13.4	168	2	F97290
12	327.5	13.4	165	2	B69059
13	327	13.4	175	2	JC7166
14	318.5	13.0	173	2	D84032
15	315	12.9	163	2	H81066
16	307.5	12.6	172	2	B64320
17	307	12.6	172	2	B56684
18	305	12.5	166	2	B97922
19	302.5	12.4	158	2	F95051
20	302	12.4	167	2	JC5165
21	290.5	11.9	171	2	C72362
22	280	11.5	190	2	AD3329
23	277	11.3	163	2	F83059
24	277	11.3	185	2	I40667
25	276.5	11.3	163	1	YC6C3H
26	275	11.3	190	2	D97604
27	275	11.3	190	2	AE2826
28	273.5	11.2	164	2	D82072
29	271.5	11.1	163	2	S15940

30	271.5	11.1	164	2	AB0517	acetoacetate synth
31	268.5	11.0	163	2	B64131	acetoacetate synth
32	268.5	11.0	163	2	B90639	acetoacetate synth
33	268.5	11.0	163	2	B85490	acetoacetate synth
34	268.5	11.0	164	2	AB0067	acetoacetate synth
35	263.5	10.8	163	2	AB1694	acetoacetate synth
36	263.5	10.8	163	2	A11322	acetoacetate synth
37	262.5	10.7	158	2	A86778	hypothetical prote
38	260.5	10.7	158	2	S35139	probable acetoact
39	255	10.4	142	2	B48648	acetylhydroxy acid
40	252.5	10.3	158	2	B84956	acetoacetate synth
41	251	10.3	232	2	T39432	acetoacetate synth
42	235.5	9.6	309	2	S19411	acetoacetate synth
43	223.5	9.1	330	2	T49352	probable acetoact
44	208.5	8.5	154	2	F81404	acetoacetate synth
45	118	4.8	2245	2	T27511	hypothetical prote

ALIGNMENTS

RESULT 1

D84725

probable acetoacetate synthase (imported) - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: D84725

R.lin: X. Kaul, S. Rounsley, S.D. Shea, T.P. Buntto, M.I. Town, C.D. Fujii, C.Y

M. Koo, H. Moffat, K.S. Cronin, L.A. Shen, M. VanAken, S.E. Umayam, L. Tallon,

euss, D. Nierman, W.C. White, O. Eisen, J.A. Salzberg, S.L. Fraser, C.M. Venter

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: AB4420; MUID:20083487; PMID:10617197

A:Accession: D84725

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-484 <STO>

A:Cross-references: GB:AE002093; NID:g4887755; PIDN:AA032291.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g31810

A:Map position: 2

Query Match	Best local Similarity	Score	2358.5;	DB 2;	Length	484;
Matches	480;	Conservative	0;	Mismatches	1;	Indels
					13;	Gaps
					2;	
QY	1	MAATSVSSPSIRCLRSACSDSSPALVSTRVSPFAKISYLSGISHRGDEMKRMGEFV	60			
DB	1	MAATSVSSPSIRCLRSACSDSSPALVSTRVSPFAKISYLSGISHRGDEMKRMGEFV	60			
QY	61	RSVDGKISDASESSAPPKSVRRKHTTSVFGDSSGMINTAGVFAARGNIESLAAG	120			
DB	61	RSVDGKISDASESSAPPKSVRRKHTTSVFGDSSGMINTAGVFAARGNIESLAAG	120			
QY	121	LNBDKALFTIIVCGTERVLAQVIEQLKLVNLKVEDISEPQVEHEMLVYNAHPESR	180			
DB	121	LNBDKALFTIIVCGTERVLAQVIEQLKLVNLKVEDISEPQVEHEMLVYNAHPESR	180			
QY	181	AETMLVDTFRRVNDIAEHALTIEVTDGPGKMAVERNLKRFQIREIVRTCK---IALR	237			
DB	181	AETMLVDTFRRVNDIAEHALTIEVTDGPGKMAVERNLKRFQIREIVRTCK---IALR	237			
QY	238	REMGATAPFWRSAASYDLKQAPVSVLRSSKKAIVPQKETSAGGVYVPEPFDDK	297			
DB	238	REMGATAPFWRSAASYDLKQAPVSVLRSSKKAIVPQKETSAGGVYVPEPFDDK	297			
QY	298	VHRIIDAHMGLTDEDTSGLRSHHTLSLVNDIPGVNITGVFAARGVNTQSLANGHAET	357			
DB	298	VHRIIDAHMGLTDEDTSGLRSHHTLSLVNDIPGVNITGVFAARGVNTQSLANGHAET	357			
QY	301	VHRIIDAHMGLTDEDTSGLRSHHTLSLVNDIPGVNITGVFAARGVNTQSLANGHAET	350			
DB	301	VHRIIDAHMGLTDEDTSGLRSHHTLSLVNDIPGVNITGVFAARGVNTQSLANGHAET	350			
QY	358	KGISRTTIVIPATDESVSRLVQOLKLVNDVHEVHDLTHLPSEERELMIKIAVNAARD	417			
DB	358	KGISRTTIVIPATDESVSRLVQOLKLVNDVHEVHDLTHLPSEERELMIKIAVNAARD	417			
QY	351	KGISRTTIVIPATDESVSRLVQOLKLVNDVHEVHDLTHLPSEERELMIKIAVNAARD	410			
DB	351	KGISRTTIVIPATDESVSRLVQOLKLVNDVHEVHDLTHLPSEERELMIKIAVNAARD	410			

[illegible]

```

Query Match          18.9%; Score 461; DB 2; Length 188;
Best Local Similarity 53.0%; Pred. No. 8.5e-24;
Matches      87;  Conservative  40;  Mismatches  37;  Indels      0;  Gaps      0;

Oy      320 HTLSLVNDIGVLNTVGVFARAGYNIQSILAVGAETKGISRTTTPVLPATDESYSKLVQ 379
      |||||:| | | | | :|:|||||:|:||||| | | | | | :|:|:| | |
Db      19 HTLSLVDEDAVGLTRINGLFAKRGFNIESLAVSGADGGVSRITMVPDGEINTLEQLTK 78

Oy      360 QLYKLVNVEHYHDLTHLPESERELMLIKIAVNAARQVLDIASIFRAKAVDSDHTITL 439
      |||||:| | | | | :|:|||||:| | | | | :| | | | | :|:|:| | |
Db      79 QLYKLVNVIKQDITETPCVERELMLVKVSNAPRAAEVIELAOFARARIADISEDPTVI 138

Oy      440 QLTGDLKMAVLQRLLEPYGICEVARTGRVLAAREGSVDSTYLR 483
      ::||| | |||||:|:| | | | |||||:| | | | | | |||||:|:| | |
Db      139 EVYDGPGRKMAVLOMLAKFGLEKARTGKIALVRESGVNTEYLK 182

RESULT 4
573151
acetoxydihydroxyacid synthase small chain - red alga (porphyra purpurea) chloroplast
C:Species: chloroplast Porphyra purpurea
C:Date: 19-Mar-1997 #sequence_revision 09-May-1997 #text_change 22-Jun-1999
C:Accession: S73151
R:Reith, M.; Munholland, J.
Plant Mol. Biol. Rep. 13, 333-335, 1995
A:Title: Complete nucleotide sequence of the Porphyra purpurea chloroplast genome.
A:Reference number: S73108
A:Accession: S73151
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-174 <REI>
A:Cross-references: EMBL:U38804; NID:g1276652; PIDN:AAC08116.1; PID:g1276696
C:Genetics: the nucleotide sequence was submitted to the EMBL Data Library, October 1995
A:Gene: 11vH
A:Genome: chloroplast
C:Superfamily: acetoalactate synthase small chain
C:Keywords: chloroplast

Query Match          17.5%; Score 427; DB 2; Length 174;
Best Local Similarity 50.6%; Pred. No. 1.4e-21;
Matches      83;  Conservative  41;  Mismatches  40;  Indels      0;  Gaps      0;

Oy      320 HTLSLVNDIGVLNTVGVFARAGYNIQSILAVGAETKGISRTTTPVLPATDESYSKLVQ 379
      |||||:| | | | | :|:|||||:|:||||| | | | | | :|:|:| | |
Db      3 HTLSLVYDDEAGVLSRISGLFAKRGFNIASLAVGPAEDIGVSRITMVPQDNRTLEQLTK 62

Oy      380 QLYKLVNVEHYHDLTHLPESERELMLIKIAVNAARQVLDIASIFRAKAVDSDHTITL 439
      |||||:| | | | | :|:|||||:| | | | | :| | | | | :|:|:| | |
Db      63 QLYKLVNLTNQQDVTNIPESVERELMLIKIQVNSQRIRALIELVKIFRANVVDIADIEDLIIV 122

Oy      440 QLTGDLKMAVLQRLLEPYGICEVARTGRVLAAREGSVDSTYLR 483
      ::||| | |||||:| | | | |||||:| | | | | | |||||:|:| | |
Db      123 EVYDGPGRKVAIEQLLTKFGIIEIKRTGKISLVRSKINTETYLK 166

RESULT 5
569464
acetoalactate synthase, small subunit (11vN) homolog - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Jun-1999
C:Accession: F69464
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Doak  

.: Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kitzness, E  

.: Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  

Nature 390, 364-370, 1997
A:Authors: Overbeek, T.; Cotton, M.D.; Spriggs, T.; Arlisch, P.; Kaine, B.P.; Sykes,  

Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archa  

.: Reference number: A69250; MUID:98049343; PMID:9369475
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Accession: F69464

```


THIS PAGE BLANK (USPTO)